

Beyond Self-Assessment: Exploring Student-Teacher Agreement on Social and Emotional Skills (SEMS) Using the Too Little/Too Much (TLTM) Scale

Word count: 13459

(Introduction – Conclusion; excluding Tables and Figures)

Jana De Smedt

Student number: 01802482

Supervisor: Prof. Dr. Bart Wille

Master's dissertation submitted in order to obtain the degree of Master of Science in Psychology
– Personnel Management and Industrial Psychology

Academic year: 2024-2025

The undersigned, Jana De Smedt, gives permission for third parties to consult this dissertation.

Abstract

Context

Teachers' social and emotional skills (SEMS) are increasingly recognised as vital for fostering positive classroom dynamics. However, research has primarily relied on teacher self-assessments to measure these competencies.

Novelty

This thesis introduces a student-based assessment using the Too Little/Too Much (TLTM) scale to explore self-other agreement (SOA) in SEMS ratings and its relation to teachers' attitudes towards feedback.

Method

Data were collected from 65 teachers and 866 secondary school students in Flanders across two time points. Exploratory factor analyses (EFA) examined the structure of SEMS ratings as measured by the TLTM scale. Correlational analyses and multilevel regression models were then used to assess agreement between teacher and student assessments. Finally, standard linear regression models investigated how perceived discrepancies relate to teachers' feedback attitudes.

Key Findings

Some TLTM subscales demonstrated acceptable reliability, while others showed variability. Agreement between teachers and students was strongest for expressive, observable behaviours such as using personal anecdotes and/or humour in class. Discrepancies between teacher self-assessments and student ratings were generally small, and overall, feedback was received positively.

Conclusion

These findings highlight the potential of student-informed feedback to enhance teacher development, particularly when it is contextualised and constructively framed. That teachers generally responded positively to feedback suggests that full alignment is not essential for it to be experienced as meaningful. The study offers preliminary evidence that nuanced, student-based assessments like the TLTM scale can enrich reflective practice and contribute to a more participatory feedback culture in education.

Nederlandse Samenvatting

Context

De sociaal-emotionele vaardigheden (SEMS) van leerkrachten worden steeds vaker erkend als essentieel voor het creëren van een positieve klasdynamiek. Toch baseert onderzoek zich doorgaans op zelfrapportage door leerkrachten om deze vaardigheden te meten.

Nieuwheid

Deze scriptie introduceert een leerlinggebaseerde evaluatie met behulp van de 'Too Little/Too Much'-schaal (TLTM) en onderzoekt de mate van overeenstemming (self-other agreement, SOA) tussen SEMS beoordelingen door leerlingen en leerkrachten. Daarnaast wordt nagegaan in hoeverre deze overeenstemming – of het gebrek daaraan – verband houdt met de houding van leerkrachten ten aanzien van ontvangen feedback.

Methode

Gegevens werden verzameld bij 65 leerkrachten en 866 leerlingen uit het secundair onderwijs in Vlaanderen, verspreid over twee meetmomenten. Via exploratieve factoranalyse (EFA) werd de structuur van de TLTM-schaal onderzocht. Vervolgens werden correlaties en multilevel regressiemodellen gebruikt om de overeenstemming tussen leerkracht- en leerlingbeoordelingen van SEMS in kaart te brengen. Tot slot werd met lineaire regressiemodellen nagegaan in welke mate waargenomen discrepanties samenhangen met de feedbackattitude van leerkrachten.

Belangrijkste Bevindingen

Sommige TLTM-subschalen vertoonden een aanvaardbare betrouwbaarheid, terwijl andere meer variabiliteit lieten zien. De overeenstemming tussen leerkrachten en leerlingen was het sterkst voor expressieve, observeerbare gedragingen zoals het gebruik van persoonlijke anekdotes en/of humor in de klas. De verschillen tussen zelfbeoordelingen en leerlingbeoordelingen bleven over het algemeen beperkt. Leerkrachten stonden bovendien overwegend positief tegenover de ontvangen feedback.

Conclusie

Deze bevindingen onderstrepen het potentieel van door leerlingen geïnformeerde feedback om de professionele ontwikkeling van leerkrachten te versterken, vooral wanneer die feedback op een constructieve en contextgebonden manier wordt aangeboden. Dat leerkrachten over het algemeen positief reageerden, suggereert dat volledige overeenstemming geen voorwaarde is voor het ervaren van feedback als zinvol. De studie

biedt voorlopige aanwijzingen dat genuanceerde, leerlinggebaseerde beoordelingen – zoals de TLTM-schaal – reflectie kunnen verdiepen en bijdragen aan een meer participatieve feedbackcultuur in het onderwijs.

Preface

This thesis marks the final step in my journey towards a Master of Science in Psychology (Personnel Management and Industrial Psychology). Writing it has been both a challenging and enriching experience, during which I have learned a great deal – not only about research, but also about perseverance and growth. I am deeply grateful to all those who supported me along the way.

I would like to thank Professor Bart Wille for his trust and for offering me the opportunity to contribute to such an engaging line of research. I am especially grateful to my supervisor, Sofie Dupré, whose guidance was invaluable throughout the process. Her thoughtful feedback and support at every stage were instrumental in shaping this thesis.

I would also like to thank Joyce Scheirlinckx for her role in the data collection, as well as Astrid for lending me her book on applied statistics in SPSS – a gesture that proved more helpful than she might have imagined. I am equally grateful to Lisa for her generous advice and encouragement, particularly during the more technical phases of the analyses.

The road to this thesis took a little longer, as I chose to combine this programme with a Master of Science in Teaching in Behavioural Sciences. Alongside that, I took on student jobs, language courses, and international engagements through the European Federation of Psychology Students' Associations (EFPSA). Still, I would not have had it any other way. I am thankful for the many opportunities these experiences offered me to grow personally and professionally. I also wish to thank those within EFPSA who believed in me and entrusted me with meaningful roles alongside my studies.

Last but not least, to my family and to Tristan (the love of my life), and his family: thank you for standing by me – for the pep talks, the practical help, the encouragement, and above all, your unwavering belief in me, even when I doubted myself. Thank you, Mum and Dad, for everything – not least for making my studies possible. And to my closest friends: your presence over the past years has been a grounding force.

To each of you – thank you, from the bottom of my heart. This achievement is mine, but it was made possible by you.

Jana De Smedt

Heusden

Monday, August 11, 2025

Table of Contents

Abstract	iii
Nederlandse Samenvatting	iv
Preface	vi
List of Tables	ix
List of Figures	x
Introduction	1
Literature Study	2
Teachers' Impact on Student Development.....	2
Social and Emotional Skills in Education.....	3
The Too Little/Too Much Scale	5
The Role of Feedback on Professional Development.....	6
Factors in the Relationship between Student Feedback and Teacher Improvement.....	8
Teachers' Feelings about Feedback.....	9
Method	11
Ethical Considerations	11
Data Collection Process.....	12
Participants	13
The Too Little/Too Much Scale: Development and Validation	14
Other Measures	16
Final Sample.....	18
Statistical Data Analysis	20
Results	23
RQ1: Structural Validity of the Too Little/Too Much Scale	23
RQ2: Self-Other Agreement in SEMS Assessment	26
RQ3: Relationship between Feedback Attitude and Discrepancy	30
Discussion	33
Reflection on Results per Research Question.....	33

Limitations of This Master's Thesis	36
Strengths and Practical Implications	38
Further Research	39
Additional Reflections: Factors in the Relationship between Student Feedback and Teacher Improvement.....	40
Conclusion	42
References	44
Appendix A: Feedback Report Examples.....	59
Appendix B: Factor Loadings.....	61
Appendix C: Scatter Plots	64
Appendix D: Dropout Analysis.....	71

List of Tables

Table 1: Descriptive Statistics for the Teachers (N = 60)	19
Table 2: Descriptive Statistics for the Students (N = 816).....	20
Table 3 TLTM Scale Validation: Internal Consistencies (α) and Explained Variances (%) Across Domains, Facets, and Final Subscales for Teachers and Students.....	24
Table 4: Model Fit and ICC Comparison Across TLTM Subscales	27
Table 5: Correlations and Regression Results for the Seven TLTM Subscales	29
Table 6: Quartile Distributions of Feedback Attitude Scores per Item and Total Scale	30
Table 7: Linear Regression Results Predicting Teacher Feedback Attitude from TLTM Discrepancy Scores	31
Table B1: TLTM Scale Validation: Factor Loadings for Teachers and Students (Appendix B)	62
Table D1: Mann–Whitney U Test Results for Teacher, Student, and Discrepancy Scores Across the Seven TLTM Subscales (Appendix D).....	71

List of Figures

Figure 1: PRISMA Flowchart of Sample Selection	15
Figure 2: Boxplots of Discrepancy Scores Across TLTM Subscales.....	32
Figure A1: Spider diagram of the current skills (SEMS FFM domains) (Appendix A)	60
Figure A2: Development opportunities for the SEMS domain 'social interaction' (Appendix A)	60
Figure C1: Scatter Plot of Teacher and Class Average 'Classroom organisation and management' Scores with Regression Line (Appendix C)	64
Figure C2: Scatter Plot of Teacher and Class Average 'Student-Centred Responsiveness' Scores with Regression Line (Appendix C).....	65
Figure C3: Scatter Plot of Teacher and Class Average 'Connecting and Caring' Scores with Regression Line (Appendix C).....	66
Figure C4: Scatter Plot of Teacher and Class Average 'Mindset About Students' Learning' Scores with Regression Line (Appendix C).....	67
Figure C5: Scatter Plot of Teacher and Class Average 'Cheerful Engagement' Scores with Regression Line (Appendix C).....	68
Figure C6: Scatter Plot of Teacher and Class Average 'Reflective and Innovative Teaching' Scores with Regression Line (Appendix C).....	69
Figure C7: Scatter Plot of Teacher and Class Average 'Emotionally Grounded Presence' Scores with Regression Line (Appendix C).....	70

Introduction

Teachers' social and emotional skills (SEMS) are increasingly recognised as crucial for fostering a supportive and inclusive learning environment. This growing recognition sets the stage for the present thesis. SEMS, encompassing capabilities like trust, empathy, and emotional self-regulation, are vital for effective teacher-student interactions. Despite extensive research and varied frameworks proposed by scholars such as Ciecuch and Strus (2021) and Primi et al. (2021), these skills are predominantly measured through self-assessment questionnaires (e.g., Akram & Zepeda, 2015; Borg & Edmett, 2019; Li et al., 2021).

This thesis introduces a novel approach by integrating student evaluations into the assessment process using the Too Little/Too Much (TLTM) scale. This scale captures the nuances of underperformance and overperformance, offering a fresh perspective on SEMS assessment. The first aim is to examine the structural validity of the TLTM scale within the context of SEMS measurement. While the theoretical domains and facets by Scheirlinckx et al. (2023) guide the initial structure, the TLTM's novel format and adapted item content call for an exploratory approach. Therefore, exploratory factor analysis (EFA) is used to uncover a more empirically grounded factor structure. This leads to the first research question: "What is the structural relationship among the items in the 'SEMS short – TLTM' rating scale when applied to the measurement of secondary school teachers' SEMS?"

Next, this thesis investigates the degree of alignment between teachers' self-assessments and student evaluations of SEMS, addressing a gap in the literature on self-other agreement (SOA). Although SOA has been studied before (e.g., Fleenor et al., 2010; Ham et al., 2015), few studies have focused on SEMS. This raises the question: "To what extent do the SEMS self-assessments of secondary school teachers and the SEMS assessments of the students in their class correspond?"

Lastly, this research explores how discrepancies between teacher and student ratings relate to teachers' attitudes towards feedback. Defined here as the emotional and belief-related responses to feedback, 'feedback attitude' is examined through the question: "To what extent, and in what direction, does the (dis)agreement between the SEMS self-assessments of secondary school teachers and the SEMS assessments of the students in their class influence the teachers' feedback attitude?" This investigation not only pioneers into the emotional consequences of SOA in educational settings but also addresses behavioural changes and feedback utilisation, areas predominantly explored in previous research (e.g., Yammarino & Atwater, 1997).

These final two questions are central to understanding the role of perception alignment in fostering a positive feedback culture within schools. By clarifying how (dis)agreement shapes teachers' responsiveness to feedback, this thesis contributes to strategies that promote feedback acceptance and professional growth. Situated in the 2024 Flemish secondary education context, this study also carries practical relevance for local educational policy and practice.

Literature Study

Teachers' Impact on Student Development

John Hattie's extensive meta-analytical research (Hattie, 2009) provides valuable insight into the many ways in which teachers and their characteristics shape student outcomes, particularly in secondary education. His synthesis of over 800 meta-analyses results in a ranked list of factors influencing student learning and achievement. Among the highest-ranking are teacher-related factors, such as instructional methods and social-emotional competencies.

A notable example is the quality of teacher-student relationships, which shows a substantial effect size of 0.72 (Hattie, 2009). This directly relates to the concept of social and emotional skills (SEMS), as it highlights the importance of fostering positive, supportive, and respectful interactions in the classroom. As Hattie (2009) notes: "developing relationships requires skill by the teacher – such as the skills of listening, empathy, caring, and having positive regard for others" (p. 118). The skills mentioned here are clear examples of SEMS. Teachers who create a secure and trusting classroom environment can significantly enhance students' motivation and engagement. Hattie (2009) thus demonstrates that teacher-student relationships are a foundational aspect of teaching, and that not only the harder skills like 'teacher subject matter knowledge' are important. In fact, teacher subject matter knowledge is shown to have an effect size of only 0.09 (Hattie, 2009, p. 109).

More recent meta-analyses further confirm the importance of teacher characteristics – such as enthusiasm, confidence, and preparation – for student learning. Schneider and Preckel (2017), for instance, highlight the strong association between classroom social interaction and student achievement. They define social interaction through five variables: the teacher's encouragement of questions and discussion, availability and helpfulness, use of open-ended questions, small-group learning, and concern or friendliness towards students. Each of these variables shows a medium to strong association with student achievement (d ranging from 0.47 to 0.77). This again indicates the importance of establishing an environment where students feel at ease to respond to questions, express their opinions, and participate in social interactions with their teacher and peers (Schneider &

Preckel, 2017). Stronge (2018) adds that teachers bring a distinctive combination of knowledge, skills, attitudes, beliefs, values, aspirations, and motivation into the classroom environment. These elements impact the classroom ambiance and teaching effectiveness, thereby affecting student outcomes.

The significance of more socially and emotionally oriented types of skills (sometimes referred to as 'non-cognitive skills', 'soft skills', 'relationship skills', 'inter- and intra-personal skills', 'character skills', 'people skills', or even 'life skills' and '21st-century skills' – underscoring their importance in contemporary life; in this thesis called SEMs) is also highlighted by theory and research related to Self-Determination Theory (SDT) (Deci & Ryan, 1985). SDT advocates that supporting students' intrinsic motivation and psychological needs – autonomy, belonging, and competence (the 'ABC's of motivation') – is crucial for success in a broad range of contexts and life domains. This includes educational contexts, where these needs underpin effective learning (Deci & Ryan, 2012). As a framework for understanding human motivation in education, SDT accentuates the role of teachers in creating environments that support these psychological needs. Teachers proficient in SEMs are more capable of fostering a learning atmosphere that meets students' needs for autonomy (by offering students choice and voice), belonging (by nurturing a sense of relatedness and connectedness within the classroom), and competence (by presenting challenges that match their skill level). By embodying skills that resonate with the ABC's, teachers can profoundly influence student motivation, engagement, and consequently, their academic achievement.

Collectively, these lines of research and theory underscore that teachers play an important role in shaping the educational and developmental outcomes of students. Drawing from these sources, it becomes evident that their impact extends well beyond the mere delivery of curriculum content. Teachers significantly impact learning by offering emotional support, boosting student motivation, and fostering a positive environment through strong teacher-student relationships, largely due to their SEMs (Allen et al., 2018). Hence, attention should be directed towards the concept of SEMs and the degree to which teachers exhibit such SEMs.

Social and Emotional Skills in Education

Within the school context, both in the research field and in practice, initial emphasis was placed on the SEMs of students, as these skills not only support learning but also prepare students for the personal and professional challenges of the 21st century (OECD, 2015). As a result, numerous initiatives have aimed to strengthen students' SEMs. For example, McNaughton et al. (2022) investigate the relationship between students' self-regulation, social skills, and academic performance in digital settings, highlighting the value

of social skills development. Similarly, Allen et al. (2018) find that emotional stability – a personality trait also considered an emotional skill since it can be developed or enhanced (Pichurin et al., 2023) – is associated with a sense of belonging at school. In addition, Miyamoto et al. (2015) note that many OECD countries include civic education in their curricula to enhance students' SEMS and better equip them for modern life.

Alongside recognising SEMS' importance for students, there is growing awareness of its critical role for teachers. A notable contribution is the literature review by Scheirlinckx et al. (2023), which synthesises different conceptualisations of teachers' SEMS. It proposes a theoretical framework grounded in the Five-Factor Model (FFM), also known as the 'Big Five' of personality (McCrae & Costa, 1987). Various SEMS of teachers are identified and categorised under the five factors (in this thesis referred to as 'domains') of this model, alongside some specific skills for the teacher's field of work: 'Professional commitment', 'Collaboration with colleagues', and 'Collaboration with parents/guardians'. Each domain includes several facets. For instance, under 'Agreeableness', the framework includes 'Trusting students', 'Proactive perspective taking', 'Connecting and caring', and 'Mindset about students' learning'. This Big Five-based structure aligns with other models of SEMS, such as the Behavioural, Emotional, and Social Skills Inventory (BESSI) (Napolitano et al., 2021; Soto et al., 2022). BESSI complements the FFM by assessing a broad range of Social, Emotional, and Behavioural (SEB) skills, thereby encompassing SEMS.

As mentioned above, teachers' SEMS influence students' academic and personal outcomes, for example, through a stronger teacher-student relationship (e.g., Hattie, 2009; Deci & Ryan, 2012; Allen et al., 2018). Interestingly, at the same time, they also influence the development of students' SEMS. For instance, Napolitano et al. (2021) underscore the instrumental role that SEB skills play across a variety of outcomes crucial for educational success. They state that "adolescence is a period of marked transitions across multiple domains, and that in order to successfully navigate those transitions, youth must use a wide range of complex, newly developing skills" (p. 5). Accordingly, they put forward adolescence as a critical period for the development and reinforcement of SEMS, advocating for targeted interventions during this period. Given that adolescents spend a significant portion of their time under the tutelage of educators within the Belgian school system, teachers emerge as pivotal figures in facilitating this developmental process. Their ability to effectively deploy SEMS can dramatically influence students' academic achievements, occupational aspirations, social connections, and overall well-being. This modelling effect is supported by Bandura's (1977) theory of observational learning, which posits that individuals learn behaviour by observing others. As students witness teachers demonstrating effective emotional regulation, empathy, and social engagement, they internalise these behaviours

and apply them in their own lives. This underscores the value of teachers' SEMs not just in direct instruction but as a pervasive influence throughout the school environment.

In conclusion, the growing recognition of SEMs within education represents a shift toward a more holistic view of learning and development. Both students' and teachers' SEMs are essential for the academic and personal growth of students. Teachers' SEMs have a significant part in creating a nurturing and supportive learning environment, and in developing the same skills in students. As such, it is important to consider how teachers' SEMs can be supported and developed. One meaningful way to do so is through feedback.

The Too Little/Too Much Scale

Before discussing the role of feedback in teachers' professional development, it is important to acknowledge that feedback is not simply a matter of good or bad performance. In Student Evaluation of Teaching (SET) measures, students may perceive teachers to exhibit certain behaviours too little (underperformance) or too much (overperformance). This phenomenon, often termed the 'too-much-of-a-good-thing (TMGT) effect' (e.g., Pierce & Aguinis, 2013; Busse et al., 2016; Vergauwe et al., 2017), occurs in educational settings as well as in workplaces. The Too Little/Too Much (TLTM) scale addresses this by providing a nuanced understanding of the balance between underperformance and overperformance, thus enriching SEMs assessments.

The development of the TLTM scale was driven by the recognition that traditional feedback mechanisms often fail to capture the full spectrum of behaviour. For example, traditional Likert-scales often treat behaviours as linear and fail to account for curvilinear (or 'inverted U-shaped') relationships between actions and their outcomes (Vergauwe et al., 2017). By assessing behaviours on a single continuum, traditional scales imply that more of a positive behaviour is always better, an idea that is also present in 'strengths-based development' approaches (Kaiser & Overfield, 2011). However, research has shown that extremes in behaviour – whether too little or too much – can negatively impact performance and effectiveness, leading to diminishing returns or even negative outcomes (e.g., Pierce & Aguinis, 2013). The TLTM scale also aims to reduce bias towards the extremities. In a traditional Likert scale, employees or students might feel pressured to rate behaviours at the high or low end of the scale, without considering the middle ground.

The TLTM scale was designed to address these limitations by allowing for a more granular assessment of behaviour, particularly in leadership and professional development contexts. It introduces a rating system that ranges from -2 (too little) to +2 (too much), with 0 representing the optimal level of performance. This format accurately identifies both deficiencies and excesses in behaviour, offering precise guidance for behavioural

adjustments. In educational contexts, this clarity enables teachers to target improvements effectively, enhancing teaching practices and, subsequently, student outcomes.

Thus, when students assess teachers' SEMs, the TLTM scale offers valuable insights into perceived behavioural balance, highlighting actionable areas for teacher development. Consequently, this thesis aims to validate the use of the TLTM scale within SEMs assessment, addressing the following research question:

RQ1: What is the structural relationship among the items in the 'SEMS short – TLTM' rating scale when applied to the measurement of secondary school teachers' SEMs?

The structure of the TLTM scale used in this study (the 'SEMS short – TLTM') is elaborated in the Method and Results sections. As introduced earlier, this thesis builds on Scheirlinckx et al. (2023), leading to the following theory-driven hypothesis:

H1: When applied to the measurement of secondary school teachers' SEMs, the items of the 'SEMS short – TLTM' rating scale are related according to the five theoretical domains and their facets, as proposed by Scheirlinckx et al. (2023).

The Role of Feedback on Professional Development

Self-reported feedback. Self-assessment is pivotal for educators' professional development, offering opportunities for self-regulation and introspection. It allows teachers to pinpoint strengths and identify improvement areas, promoting continuous professional growth (Ross & Bruce, 2007). As noted by Eva and Regehr (2005), self-assessment fulfils two key functions: (1) identifying gaps in knowledge and skills, and (2) reinforcing existing expertise and abilities. For instance, when teachers accurately identify areas where their knowledge is lacking, they are more inclined to seek out resources to close these gaps, thereby enhancing their teaching practices. On the other hand, awareness of their strengths enables teachers to maintain confidence in applying effective strategies.

The success of self-assessment in fostering professional development is largely dependent on its accuracy. Several researchers (e.g., Aesaert et al., 2017) have suggested that self-assessments reflect individuals' self-perceived competencies, which may not align with their actual abilities. When teachers have inaccurate self-perceptions, their capacity to undertake meaningful improvement can be impaired. Ernst et al. (2023) observe that many teacher candidates tend to be either overconfident or underconfident about their knowledge, often lacking awareness of their own strengths and weaknesses. Also, studies highlight the difficulty of achieving objectivity in self-assessments, with many educators tending to overestimate their skills in certain areas while underestimating them in others (e.g., Montgomery & Baker, 2007; Ernst et al., 2023). This imbalance indicates that, while self-

assessment offers important self-reflection, it often needs to be complemented by external feedback (from students, for instance) to ensure a more balanced perspective (Stalmeijer et al., 2009).

Student-generated feedback. The significance of student feedback in educational settings is extensively documented, underscoring its multifaceted role in enhancing teaching quality, promoting professional growth, and fostering faculty development (e.g., Ferguson et al., 2011; Husain & Khan, 2016; Karm et al., 2022; Cornes et al., 2023). It allows teachers to make real-time adjustments in their teaching practices, thereby enhancing the learning experience for both current and future classes. Emphasising this to the students can motivate them to provide honest anonymous feedback to their teachers (Caulfield, 2007).

Moreover, student feedback is linked to improved academic outcomes, suggesting it's not only a tool for evaluating teaching but also a mechanism for boosting student performance and learning gains (e.g., Caulfield, 2007; Yu, 2016). According to Caulfield (2007), the implementation of student feedback can lead to improved grades, signifying enhanced learning outcomes – although students themselves do not seem to see this relationship. Yu (2016) expands on this perspective by highlighting the critical role of feedback in the refinement of teaching methodologies. This study emphasises that feedback, especially when utilised formatively, contributes to substantial learning gains across diverse educational settings and levels of education (diploma and degree levels), and content areas.

Self-other agreement. In the context of professional development, understanding the dynamics of self-other agreement (SOA) plays a pivotal role, particularly in educational settings where teacher-student feedback is crucial for improving teaching practices. SOA refers to the congruence between an individual's self-assessment and the assessments provided by others, such as peers, subordinates, or superiors (Fleenor et al., 2010). This concept is highly relevant when examining the extent to which teachers' perceptions of their own effectiveness align with the perceptions of their students (Ham et al., 2015).

The process of obtaining feedback from multiple sources, often through 360-degree feedback mechanisms, is a common approach to assess SOA (e.g., Atwater & Yammarino, 1992; Lee & Carpenter, 2018). This method has been shown to enhance the validity of performance evaluations by providing a holistic perspective on an individual's competencies (e.g., Atwater et al., 1997; Fleenor et al., 2010). In educational settings, student feedback on teachers' social-emotional and leadership skills can serve as a valuable indicator of how accurately teachers perceive their own strengths and weaknesses.

Bergner et al. (2016) state that “agreement has been linked to positive workplace outcomes” (p. 335). Indeed, research shows that SOA can have significant implications for

professional development. For example, a lack of congruence between self-assessments and external evaluations often indicates discrepancies in self-awareness. These discrepancies, in turn, influence how individuals respond to feedback and their willingness to engage in developmental activities (Atwater & Yammarino, 1992). MacKie (2015) furthermore demonstrates that managers adjust their self-perception based on the feedback received. This results in managers who initially over-rated themselves reducing their self-ratings over time, resulting in more SOA. Viewing the feedback from their students alongside their own self-ratings may thus lead teachers to develop greater self-awareness. This, in turn, might make them more likely to accept constructive feedback from students and adjust their self-perceptions accordingly.

It becomes clear that numerous studies have investigated SOA. However, none have specifically focused on teachers' SEMS. Therefore, a second research question explored in this study is:

RQ2: To what extent do the SEMS self-assessments of secondary school teachers and the SEMS assessments of the students in their class correspond?

Factors in the Relationship between Student Feedback and Teacher Improvement

As indicated above, student-generated feedback has many advantages and strengths. However, the scientific literature also reveals many success factors that should be taken into account. Of course, it is not within the scope of this master's thesis to discuss them all in detail, but some are addressed here.

Specificity and accuracy. Feedback should be detailed, targeting specific behaviours or results to enhance accuracy and effectiveness (e.g., Aguinis et al., 2012; Moralejo et al., 2019; Cornes et al., 2023). Providing specific evaluation criteria to the person(s) giving feedback (in this case, students) can help (Yu, 2016).

Timing. The timing of feedback is crucial. Immediate feedback aligned with specific achievement goals can significantly improve performance (Anseel et al., 2010). Within the school context, moving the feedback exercise to the middle or the second half of the semester allows students to better adapt to lecturers' teaching styles, thereby providing more accurate and helpful feedback (Yu, 2016; Karm et al., 2022).

Individual differences. Individual differences, particularly the Big Five personality traits – extraversion, openness, agreeableness, emotional stability, and conscientiousness – moderate how negative feedback affects job performance. For example, employees (teachers) with higher conscientiousness may more readily acknowledge performance gaps,

motivating them to improve or alter behaviours to align better with organisational (school) goals (Guo et al., 2017).

Perception and engagement. The effectiveness of student feedback in teacher improvement hinges significantly on both the teachers' perception of and their engagement with the feedback. Initially, teachers' perception of the feedback – as accurate, reliable, and valuable – shapes their receptiveness. Studies show that a substantial proportion of faculty members value student evaluations, considering them reflective of their teaching performance, thus fostering a positive perception that encourages deeper engagement (Murray, 1997; Husain & Khan, 2016; Hammer et al., 2018).

However, the benefits of positive perceptions can only be realised through active engagement. Teachers must not only appreciate the feedback but also commit to processing and implementing it. Lack of engagement, regardless of perception, can lead to negative outcomes such as reduced subordinate (student) satisfaction and poorer (educational) results (Atwater & Brett, 2006). Active engagement with feedback is linked to significant achievement gains (Brooks et al., 2021).

Teachers' Feelings about Feedback

Feedback attitude and feedback orientation. Above, the importance of the feedback receiver's perception of and engagement with the feedback was already mentioned. A central concept in this context is 'feedback attitude', which refers to the feelings and beliefs a person holds about received feedback. Sometimes labelled 'feedback experience' (e.g., Gan, 2020; Zong et al., 2021), it forms a key part of the broader construct 'feedback orientation' (e.g., London & Smither, 2002; Linderbaum & Levy, 2010; Yang et al., 2023).

Rooted in organisational psychology, feedback orientation reflects an individual's general openness to and engagement with feedback as a tool for self-improvement (London & Smither, 2002; Linderbaum & Levy, 2010; Yang et al., 2023). It includes several dimensions: comfort or satisfaction with feedback; receptivity, including acceptance and agreement; and a mindful approach to interpreting and acting on feedback. Individuals with high feedback orientation typically value feedback, are attuned to environmental cues, and take responsibility for using the information constructively (Linderbaum & Levy, 2010). In contrast, a defensive or distrustful stance may hinder the meaningful integration of feedback (London & Smither, 2002).

Recent research in educational settings further extends this concept by identifying more specific facets, including feedback utility (seeing feedback as beneficial), feedback self-

efficacy (confidence in interpreting feedback), feedback social-awareness (maintaining positive feedback-provider relationships), and feedback accountability (a sense of responsibility to act on feedback) (Yang et al., 2023). While often considered a relatively stable trait, feedback orientation may evolve over time through development and environmental change (Linderbaum & Levy, 2010).

Teachers' attitudes towards (student) feedback are shaped by prior experiences and broader cultural norms. For instance, Palk (2018) finds that positive perceptions of student feedback are more common among teachers who view such feedback as reliable and valuable. Similarly, Bramley et al. (2020) highlight that teachers are more receptive to feedback when it is perceived as fair and caring, seeing it as a useful tool for self-reflection and improvement. In contrast, feedback seen as overly critical or lacking in support may deter engagement and hinder teaching adjustments.

Taken together, these studies demonstrate that the role of feedback in teacher development is multifaceted, involving both cognitive and affective processes captured in the notion of 'feedback attitude'. A supportive institutional feedback culture may meaningfully influence this attitude in teachers, ultimately benefiting student learning outcomes (e.g., London & Smither, 2002; Yang et al., 2023).

Self-other agreement and feedback attitude. In broader organisational research, SOA has been associated with important outcomes. Yammarino and Atwater (1997) found that congruence between self-assessments and others' evaluations tends to enhance both individual and organisational performance. Misalignment, by contrast, often elicits defensive reactions and hampers learning. Mackie (2015), however, observed that managers who overestimated themselves were more likely to revise their self-view in the face of contradictory feedback, suggesting a corrective effect of dissonant input.

Nielsen et al. (2021) further nuance these findings by distinguishing between 'in-agreement good leaders' and 'in-agreement poor leaders'. They suggest that poor leaders who recognise their own shortcomings may lack the confidence or drive to improve, thereby reflecting the Golem effect (Babad et al., 1982), where low mutual expectations constrain development. Similarly, in educational contexts, Devos et al. (2013) found that Flemish school leaders responded differently to feedback depending on their self-assessment and its alignment with subordinate ratings. Over-estimators tended to rationalise negative input to preserve self-esteem, potentially resulting in stagnation, while under-estimators more often interpreted discrepancy as affirmation of untapped potential.

These findings are echoed in other studies linking SOA to feedback reactions (e.g., Atwater & Yammarino, 1992; Tsui & Ashford, 1994; Sosik & Godshalk, 2004; Berson & Sosik,

2007). Collectively, they suggest that the direction and extent of (dis)agreement in performance evaluations can shape how individuals interpret, value, and respond to feedback – including whether it is internalised or dismissed.

Although feedback attitude has been linked to general perceptions of usefulness and fairness, little research has examined how SOA specifically shapes this attitude. Most existing studies focus on behavioural outcomes following feedback, whereas much less is known about the emotional and belief-based responses to feedback misalignment. This underscores the importance of investigating the congruence between teachers' self-assessments and student evaluations, and how such (dis)agreement may shape their feedback attitude. Hence the third research question:

RQ3: To what extent, and in what direction, does the (dis)agreement between the SEMS self-assessments of secondary school teachers and the SEMS assessments of the students in their class influence the teachers' feedback attitude?

Method

This master's thesis is part of a broader series of studies exploring the social and emotional skills (SEMS) of teachers, extending the research conducted by Scheirlinckx et al. (2023). The data were therefore collected by six Master's students. This also increased the chance of a sufficiently large sample, since participation was voluntarily and the teacher population in Flanders is overworked and over-questioned due to a big teacher shortage (e.g., European Commission, 2022; Whitehead, 2024). The questionnaires used were drawn up by the thesis supervisors with all the agreed variables included. This master's thesis only describes the variables that are relevant to this research.

Note that all data used in this research is original and collected specifically for the purposes of this master's thesis. No data were acquired from external, pre-existing databases or third-party sources. All analyses were conducted using the IBM Statistical Package for the Social Sciences (SPSS), version 29.0.2.0 (20).

Ethical Considerations

The study was conducted in compliance with ethical guidelines and was approved by the relevant ethics committee (reference: 2023-142). As shown in the data collection procedure (below), all participants (teachers and students alike) received clear information about the study's purpose, procedures, and confidentiality through information letters. Signed informed consent was obtained from teachers, students, and the students' parents before participation. For data collection, unique participant codes were used to ensure anonymity. Teachers and students were informed that their responses would be anonymised and used

solely for research purposes. Participation was entirely voluntary, and individuals could withdraw at any point without consequences. To reduce the burden on participants, clear instructions were provided to both teachers and students, facilitating a more efficient data collection process.

Data Collection Process

This study employed a longitudinal, quantitative research design using survey-based self- and other-assessments of SEMS at multiple time points. Given the study's focus on psychometric validation and the quantification of self-other agreement (SOA), a quantitative approach was most suitable to meet the research objectives.

For teachers, all data collection was conducted online. In the first phase (T1), between January and February 2024, participating teachers were recruited through online platforms, school visits, and personal networks. Each teacher received an information letter detailing the study's purpose, procedures, and ethical considerations, along with an informed consent form. After providing informed consent, teachers were assigned a unique code to ensure anonymity throughout the study. They then completed an online survey administered via Qualtrics. This survey captured teacher characteristics, including demographic information such as age, gender, and seniority, as well as work-related variables like job satisfaction, turnover intention, and weekly working hours. Teachers also provided class-specific details, such as the subject taught, hours of weekly contact with the class, class size, academic year, and school. Additionally, teachers self-assessed their SEMS using the 'SEMS complete' and 'SEMS short – TLTM' (self-rating) scales. After completing the survey, teachers were provided with pre-prepared envelopes containing an instruction document, information letters and informed consent forms for students' parents, and student questionnaires. Teachers distributed the materials to the selected class, collected signed parental consent forms, and administered the student questionnaires – which also had a unique code for each student written on top of it – under their supervision. Once completed, the student questionnaires were returned to the researcher, and the data were securely entered into a database for analysis.

In March 2024, teachers entered the second phase of the study (T2), during which they received a personalised feedback report via email (for some examples from such a report, see Appendix A). This report was accompanied by an explanatory video to help interpret the findings. Teachers were also asked to set development goals based on their feedback, complete a short survey about their motivation for improvement, and answer an open-ended question about their plans for professional development.

In May 2024, teachers participated in the third phase of data collection (T3). They once again completed the 'SEMS complete' and 'SEMS short – TLTM' (self-rating) scales via Qualtrics. Updated student questionnaires were provided to teachers, who distributed and supervised their completion in the same class as at T1. The completed questionnaires were then returned to the researcher.

In June 2024, the final phase of data collection (T4) involved providing teachers with a second personalised feedback report, again accompanied by an explanatory video.

For students, data collection was conducted exclusively on paper. At T1, between January and February 2024, teachers distributed information letters and informed consent forms to students' parents. Once signed consent was obtained, students completed a paper-based survey in class. This survey included demographic information such as age, gender, academic year, and school, as well as measures of emotional well-being and academic performance. Students also assessed their teacher's SEMS using the 'SEMS short – TLTM' (other-rating) scale. In May 2024 (T3), students completed a second round of surveys, which included a demographic verification and a reassessment of their teacher's SEMS using the same 'SEMS short – TLTM' scale.

Note that T3 and T4 are not discussed further in the context of this study, as the stated research questions only require T1 and T2 to be explored. This thesis furthermore only uses the 'SEMS short – TLTM' rating scales, not the 'SEMS complete'. In what follows, the 'SEMS short - TLTM' scales (31 items) are often referred to as 'TLTM' scales for simplicity.

Participants

Initial recruitment and sample reduction. At T1, 65 teachers with their classes were initially recruited to participate in the study. Out of these, 866 students filled out the questionnaire as instructed. Early exclusions were made by the thesis supervisors for students who explicitly misinterpreted (renamed) the scale on their questionnaires, rendering their responses unsuitable for analysis. Additionally, one physical education teacher (and their class of 10 students) was removed due to not meeting the inclusion criteria outlined below.

Inclusion criteria. A number of agreements had been made prior to approaching teachers. The participating teachers had to teach in secondary education, but not in special education (for students requiring adapted learning environments due to learning difficulties or disabilities). This decision was informed by the specialised nature of pedagogical strategies employed in special education settings, which may not align with the standardised

assessments used in this research. Teachers could work in any stage and grade and teach any subject. Only physical education teachers were excluded from the joint sample because they teach in a very different way and outside classrooms and therefore most likely use different SEMs. The teachers had to teach in a Flemish school and not in Wallonia. It did not matter how long a teacher had been in the classroom (seniority), but interns were excluded.

There were also some inclusion criteria regarding the choice of participating classes, which the teachers received in an information letter. The teachers were supposed to choose the class that they saw the most number of hours per week. If the teachers taught multiple classes for the same number of hours, they were asked to choose the class of which they were the 'head teacher' or 'tutor'. (In the Belgian school system, a teacher is often responsible for a specific class, overseeing the students' overall well-being, academic progress, and serving as a primary contact for parents. They coordinate with other subject teachers about students' performance and needs.) Each class was also only allowed to participate once, so if multiple teachers from the same school participated, this had to be taken into account. Note that a lesson 'hour' in Flanders corresponds to 50 (not 60) minutes.

The Too Little/Too Much Scale: Development and Validation

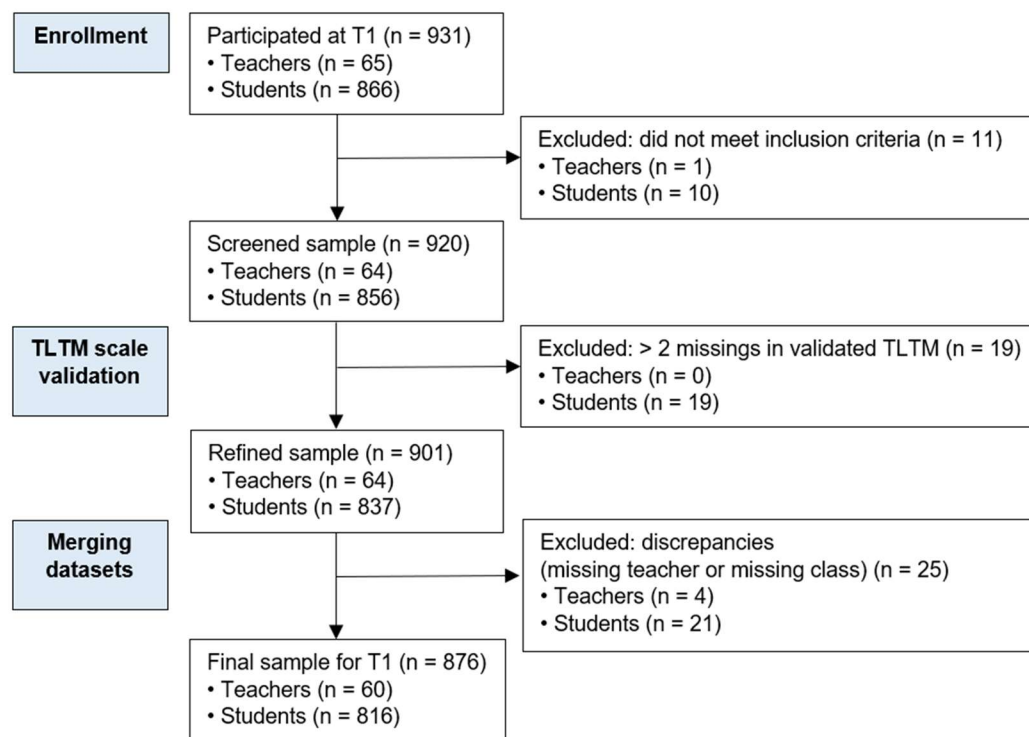
Data cleaning. Before the TLTM scale validation (within the context of SEMs measurement; encompassing factor and reliability analysis), there were 31 items. Following the validation process, 26 items remained. All 64 teachers completed all 26 items, but a significant number of students (112 of the 866) skipped at least one item. Since factor analysis requires complete data, these cases either needed to have the missing data removed or replaced. Removing all 112 students with at least one missing item would have resulted in a loss of 12.93% of the student data. Moreover, this action would lead to distortions, as the missing data did not appear to be completely random. For instance, one student skipped all 26 items! The distribution of missing values was as follows: 75 students had one missing, 18 had two, 8 had three, 4 had four, and 7 students had five or more missing responses. Five missing values correspond to not completing 19.23% of the scale, which is substantial. Even with three missing values, more than 10% of the scale was incomplete. A threshold of 10.00% was therefore set: one or two missing values were deemed acceptable, and 19 students were excluded for skipping more than two items. Mean imputation was applied for the remaining cases, using the class average for each missing item.

Merging datasets. Prior to merging the teachers and students datasets, another 4 teachers and 21 students were removed because it became apparent that they were missing

in the other dataset. Figure 1 provides an overview of all the steps taken to get from the initial to the final sample.

Figure 1

PRISMA Flowchart of Sample Selection



Note. T1 = Time 1 (the first phase of data collection); TLTM = Too Little/Too Much.

Structural validity (RQ1). To address RQ1, the underlying structure of the TLTM scale was examined to determine how its items interrelate and whether they reflect the five theoretical domains identified by Scheirlinckx et al. (2023) using traditional Likert-type ratings. Given that this study employed a newly developed response format and adapted items, it was primarily exploratory in nature.

Although the theoretical domains and facets guided the initial structure, it could not be assumed that this structure would transfer to the TLTM format. Therefore, an exploratory factor analysis (EFA) was conducted to identify a more empirically supported structure and to further assess construct validity. The validation was based on a preliminary dataset (not the final sample; see Figure 1). As the TLTM scale does not contain reverse-coded items, no reversals were required prior to analysis.

EFA was performed separately for the teacher and student data, focusing on the 31 items that were identical across both versions. The teacher data were used as the basis for

initial scale development due to their completeness and alignment with the self-assessment format. A single-factor solution explained only 13.21% of the variance in the teacher data and 17.82% in the student data, indicating that a unidimensional structure did not adequately capture the complexity of the data. Therefore, a multi-factor solution was pursued.

The extraction method was Principal Component Analysis (PCA) with Varimax rotation, chosen to identify uncorrelated components. However, in each analysis – whether conducted at the level of the full scale, individual domains, facets, or subscales – only a single component was extracted, making rotation impossible. All cases with missing values were excluded using listwise deletion. Subscale development was guided by theoretical expectations and empirical criteria such as internal consistency, percentage of variance explained, and factor loadings. Factors were retained if they accounted for a meaningful proportion of variance and yielded conceptually and statistically coherent item groupings. Following Streiner (1994), explained variances of $\geq 50\%$ were considered acceptable benchmarks for factor retention in applied psychological scale development. Deviations from this threshold were permitted when theoretical interpretability and internal consistency supported retention.

Internal consistency was assessed using Cronbach's alpha (α). Although α assumes tau-equivalence and may underestimate reliability when this assumption is violated (Cortina, 1993; Dunn et al., 2014), it remains the most commonly reported internal consistency index in SEMS and similar research (Dunn et al., 2014). Subscales were retained if they showed acceptable internal consistency ($\alpha \geq .60$) and demonstrated conceptual coherence. The factor structure derived from the teacher data was subsequently applied to the student dataset to assess cross-group consistency.

The resulting subscales (see Results) were evaluated based on both thematic coherence and internal consistency. Full reliability coefficients and explained variances for both groups are presented in the Results section. Factor loadings are reported in Appendix B.

Other Measures

In addition to the SEMS items (complete and short questionnaires), the teacher and student questionnaires used contained several other questions. Most of these additional variables were descriptive. Below only the variables that are relevant for and used in this master's thesis are discussed.

Personal, teaching, and classroom characteristics. At T1, teachers were asked to share their gender, age, tenure, the school they work for, the number of students in their

chosen class, the grade of the chosen class (from 1st to 7th), the 'programme of study' of the chosen class ('ASO' = General Education, 'TSO' = Technical and Vocational Education, 'BSO' = Vocational Education and Training, 'KSO' = Arts Education, or 'Other'), the subject(s) they taught to the chosen class, the number of hours per week they taught the chosen class, and the number of hours per week they taught in general. The number of students in the participating class as reported by the teacher, was the total number of students in that class, regardless of who actually completed the questionnaire. The subjects teachers could choose from were 'Languages (Dutch, French, German, English, Spanish, etc.)', 'Mathematics and statistics', 'Sciences (physics, chemistry, biology, etc.)', 'Human sciences (history, geography, philosophy, psychology, pedagogy, economics, etc.)', 'Art and culture (music, drama, visual arts, P.O., etc.)', 'Technical and practice-oriented subjects (computer science/electronics, fashion/clothing, cooking, wood/metal working, agriculture, etc.)', 'L.O. and sports', 'Social education (religion, ethics, philosophy of life, citizenship)', 'Classical subjects (Latin, Greek)', and 'Other'.

Students were asked at T1 to share their gender, age, school, and grade as well. They were also asked to share their scores (in percent) for the teacher's subject and in total (for all subjects combined) on the last report. The students' scores can be interesting in an evaluation context such as this: if the student consistently scored poorly in mathematics and was allowed to assess the mathematics teacher anonymously, he or she may have been very negative towards the mathematics teacher.

Feedback attitude. At T2, teachers were asked to evaluate their 'feedback experience' or 'feedback attitude', the central construct in RQ3. This construct was operationalised using five items, originally formulated in English by Anseel et al. (2010). These items were translated into Dutch following a forward translation approach, with semantic equivalence checked by the supervising team. The (English) items were: "The feedback I received was an accurate evaluation of my performance", "I believe the feedback received was correct", "The feedback I received helped me learn how I can improve my performance", "Overall, I was satisfied with the feedback I received", and "I felt quite good with the performance feedback I received". Responses were recorded on a five-point Likert scale ranging from 1 (Totally disagree) to 5 (Totally agree).

These items can be conceptually linked to the dimensions of feedback orientation discussed in the literature. The items on perceived accuracy and correctness reflect receptivity (acceptance and agreement), the items on satisfaction and emotional reaction align with comfort (satisfaction), and the item on learning potential captures a mindful

approach to interpreting and acting on feedback (cf. London & Smither, 2002; Linderbaum & Levy, 2010; Yang et al., 2023).

However, given the limited number of items and the absence of multiple indicators per dimension, no subscales were constructed. Instead, a global feedback attitude score was calculated as the average of the five items. This approach follows Anseel et al. (2010), whose EFA demonstrated that all items converge onto a single factor with satisfactory internal consistency ($\alpha = .82$). Each item has also been used in prior feedback research (Korsgaard, 1996; Keeping & Levy, 2000; Tonidandel et al., 2002; Kinicki et al., 2004).

In the present study, the scale demonstrated good internal consistency as well ($\alpha = .79$). While one item ("The feedback I received helped me learn how I can improve my performance") slightly lowered the α if removed, it was retained due to its conceptual relevance.

Final Sample

Teachers. As indicated in Figure 1, the final sample at T1 consisted of 60 teachers. Their descriptive statistics are shown in Table 1. As can be seen there, the gender distribution indicated a predominant identification as female (73.33%). Regarding age, a significant proportion of the sample (48.33%) ranged from 25 to 42 years of age, highlighting a mid-career teacher demographic within the cohort. Experience levels among the teachers varied, with half of the sample possessing between 2 and 18 years of tenure. This distribution points to a moderate level of professional experience within the group. Class sizes reported ranged from a minimum of 6 to a maximum of 26 students, with the most common sizes being 20 to 21 students, each reported by seven teachers.

The most frequently reported current classes were in their 6th year (26.67%) and labelled as 'ASO' (53.33%). Note that this study was set against ongoing reforms in Flanders' secondary education, initiated in September 2019. These reforms have significantly changed educational structures, introducing new categories that differ from previous classifications like ASO, TSO, KSO, and BSO, referenced in the surveys used. As a result, about 20% of the teachers were unable to categorise their classes under the old programmes.

Many teachers taught more than one subject. Language subjects were most common, making up 36.67% of courses, followed by Human Sciences (25.00%), and Mathematics (13.33%). Notably, no Classical languages such as Latin or Greek were taught.

The interaction frequency between teachers and their classes also varied, with the most common duration being 4 'hours' per week, as reported by 28.33% of the sample. The overall range of teaching hours per week for specific classes spanned from 1 to 12 ($M =$

4.65, $SD = 2.23$). When considering the general teaching workload, most teachers reported teaching between 20 and 22 'hours' per week, with overall reported hours ranging from 4 to 26.

Table 1

Descriptive Statistics for the Teachers (N = 60)

Variable	Category	n	%
Gender	Male	16.00	26.67
	Female	44.00	73.33
	Unknown	0.00	0.00
Age	[M, SD]	43.88	9.94
Tenure	[M, SD]	19.43	10.01
Class size	[M, SD]	17.47	4.55
Stage	1st	16.00	26.67
	2nd	15.00	25.00
	3rd*	29.00	48.33
Programme of study	ASO	32.00	53.33
	TSO	9.00	15.00
	BSO	7.00	11.67
	KSO	0.00	0.00
	Unknown	12.00	20.00
Course**	Languages	22.00	36.67
	Mathematics	8.00	13.33
	Sciences	5.00	8.33
	Human sciences	15.00	25.00
	Art	2.00	3.33
	Technical	3.00	5.00
	Social education	6.00	10.00
	Classical	0.00	0.00
	Others	7.00	11.67
Hours/week in front of class***	1 – 4	33.00	55.00
	5 – 8	25.00	41.66
	9 – 12	2.00	3.33
Hours/week total***	[M, SD]	19.08	4.09

Note. M = mean; SD = standard deviation; ASO = General

Education; TSO = Technical and Vocational Education; BSO =

Vocational Education and Training; KSO = Arts Education

*Also includes the 7th grade.

**The total may exceed 100%, since teachers can teach multiple subjects.

***1 hour = 50 minutes.

Table 2*Descriptive Statistics for the Students (N = 816)*

Variable	Category	n	%
Gender	Girl	479.00	58.70
	Boy	326.00	39.95
	Other	11.00	1.35
Age	[M, SD]	15.37	1.93
Grade	1st	72.00	8.82
	2nd	138.00	16.91
	3rd	94.00	11.52
	4th	90.00	11.03
	5th	161.00	19.73
	6th	250.00	30.64
	7th	11.00	1.35

Note. M = mean; SD = standard deviation.

Students. The final sample for T1 consisted of 816 students (see Figure 1). Analysis of the descriptive statistics (see Table 2) revealed discrepancies in the distribution of grades between teacher and student reports, which prompted a recoding. Responses were adjusted in SPSS where necessary, aligning student data either with the majority of classmates or with the teacher's report, depending on the case. The resulting distribution was skewed towards upper secondary levels, with nearly a third of students in the 6th grade. The 7th and 1st grades were least represented. Most students fell within the third educational stage, which includes grades 5 to 7.

In terms of gender, almost 60.00% of the students identified as girls, with a small but important representation (1.35%) opting for 'Other'. This highlights a diversity in gender identification among the students that was not prominent among the teachers. The ages of students ranged from 11 to 20 years, with a significant concentration among older adolescents: 27.08% were 17 years old and 18.01% were 16 years old. This age distribution reflects the high number of 6th graders in the final sample.

Statistical Data Analysis

Note that RQ1, which examined the structural validity of the TLTM rating scale, was addressed in the 'Structural validity' section above.

RQ2: Self-other agreement in SEMS assessment. To examine alignment between teacher self-assessments and student evaluations of SEMS, analyses were conducted using both aggregated (wide format) and non-aggregated (long format) student data.

Before aggregation, intraclass correlation coefficients (ICCs) were calculated to evaluate the appropriateness of class-level aggregation. These ICCs quantify the proportion of variance in student ratings attributable to the teacher/class level. Higher ICCs suggest stronger within-class agreement, supporting aggregation. Adjusted ICCs were derived from the random-intercept-only model (Model 2), following the recommendation that ICCs should be derived from models estimating group-level intercept variance only (Kreft & De Leeuw, 1998; Hox, 2010; Hox et al., 2018). The values ranged from .123 to .267, with a mean of .172 across the seven subscales. For educational research, this indicates meaningful within-class homogeneity (Hox, 2010; Hox et al., 2018).

In the wide format, each row represents a teacher, combining their characteristics with the aggregated TLTM subscale scores of their students. Aggregation was performed in SPSS by clustering students within classes based on the teacher code and computing weighted means per subscale. This format is advantageous for analyses that require a holistic view of the teacher's data alongside their class's aggregated assessments. In the long format, the dataset is structured such that each row constitutes a unique instance, specifically associating a student's evaluation with their teacher's characteristics. Teacher characteristics are thus repeated across multiple entries: one for each student. This detailed, expanded format supports analyses focused on individual-level variations and interactions within the classroom, highlighting the diversity of student perceptions regarding their teacher's SEMs.

The wide format was used to calculate Pearson correlations (r) between teacher and student (class) scores, giving a first indication of the level of agreement between them (for a visual representation, see Appendix C). The long format was used as the basis for linear mixed models. For each TLTM subscale, four models were estimated. Model 1 was a simple regression without clustering. Model 2 introduced a random intercept for each class to account for baseline differences. Model 3 further added a random slope, allowing the teacher-student score association to vary between classes. Model 4 extended either Model 2 or 3 – based on which had the better fit – by including covariates. These comprised teacher-level variables (gender, tenure, class size, and class stage – stage 3 also included grade 7) and a student-level variable (gender).

Model selection was based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, with preference given to models with lower values. In cases where the AIC indicated superior fit but the BIC value was slightly higher, AIC was prioritised for three reasons: (1) its focus on predictive accuracy aligns with the applied aims of this research; (2) its less severe penalty for model complexity supports detecting nuanced

effects; and (3) its better performance with large datasets enhances robustness (Chakrabarti & Ghosh, 2011; Diaries, 2024; Machine Learning Quick Reads, 2024).

RQ3: Relationship between feedback attitude and discrepancy. To investigate RQ3, several analytical steps were undertaken. All analyses were based on the merged dataset from T2, which contained complete data for 42 teachers. Eighteen of the original 60 teachers (final sample at T1) thus did not complete the measures at T2, yielding a dropout rate of 30%.

Variable preparation. First, seven teacher TLTM subscale scores, seven corresponding student TLTM subscale averages (aggregated at class level), and seven discrepancy scores were imported into the T2 dataset. Discrepancy scores were computed by subtracting the student class average from the teacher's self-rating per subscale. Thus, positive scores indicate that the teacher rated themselves higher on a subscale than the class average, while negative scores reflect the reverse. Importantly, the same discrepancy value may result from different rating combinations – e.g., a discrepancy of +2 could reflect a teacher rating of +2 (Too much) and a student average of 0 (Just right), or a teacher rating of 0 and a student average of -2 (Too little).

Control variables were also included, consistent with the regression model used for RQ2 (i.e., gender, tenure, class size, and stage of the class at the teacher level, but not the gender of the students). No missing values were present in the merged dataset.

Preliminary analyses and model selection. Given the relatively small sample size ($n = 42$), the number of predictors in the regression model had to be limited to avoid overfitting. As a result, interaction effects and item-level analyses were deliberately avoided. Although common guidelines suggest at least 10 observations per predictor (e.g., Harris, 1985), this threshold was not fully met. The results should be interpreted with caution due to the increased risk of both Type I and Type II errors (Babyak, 2004; VanVoorhis & Morgan, 2007).

Although the dataset included longitudinal data, the analysis does not allow for causal inference. Rather, the aim was to explore whether systematic relationships exist between perceived alignment and teachers' emotional and cognitive responses to the feedback received.

Prior to conducting the regression analyses, assumptions of linear regression were assessed. Inspection of the correlation matrix confirmed that no pairwise correlations among the discrepancy variables exceeded .50, with the highest correlation (.476) found between discrepancy in 'Emotionally grounded presence' and discrepancy in 'Cheerful engagement'.

Furthermore, all tolerance values exceeded .40, with the lowest value (.544) observed for 'Reflective and innovative teaching', suggesting that collinearity was not a concern. According to Menard (1995), low tolerance values indicate potential multicollinearity, and values substantially above zero generally do not pose problems.

Analysis strategy. As the dependent variable – feedback attitude – was continuous, linear regression analysis was deemed appropriate.

Two regression models were estimated:

1. Model A (Main model): included only the seven discrepancy scores as predictors.
2. Model B (Extended model): added the four control variables to Model A.

Model performance was assessed using Adjusted R-squared values, the AIC, and the BIC, which offers a principled approach to comparing model quality while penalising unnecessary complexity (Chakrabarti & Ghosh, 2011).

Finally, all analyses were preceded by a dropout analysis comparing teachers who remained in the study at T2 to those who dropped out after T1. To examine differences in T1 scores (teacher self-assessments, student assessments, and discrepancy scores) between these two groups, the normality of each subscale was assessed using Shapiro-Wilk tests, histograms, Q-Q plots, and boxplots, separately for teachers who dropped out and those who remained. As normality assumptions were not consistently met, t-tests were not appropriate, and Mann-Whitney U tests were used instead (Gibbons & Chakraborti, 2011).

Seven Mann-Whitney U tests were conducted – one for each SEMS subscale – prompting the application of a Bonferroni correction to control for the inflated risk of Type I error. The corrected α level was set at .007 (i.e., $\alpha = .05 / 7$ tests) per comparison set (Bland & Altman, 1995; Armstrong, 2014). An overview of the test results for each SEMS subscale is presented in Appendix D.

Results

RQ1: Structural Validity of the Too Little/Too Much Scale

The exploratory factor analysis (EFA) led to the identification of seven subscales (see Table 3), several of which partially reflected the domains and facets proposed by Scheirlinckx et al. (2023). These subscales generally explained at least 50% of the variance, aligning with recommended thresholds for multi-dimensional scale development (Streiner, 1994).

Table 3

TLTM Scale Validation: Internal Consistencies (α) and Explained Variances (%) Across Domains, Facets, and Final Subscales for Teachers and Students

TLTM item (with item number)	PER DOMAIN AND FACET		FINAL SUBSCALES	
	Teachers α ; Explained Variance	Students α ; Explained Variance	Teachers α ; Explained Variance	Students α ; Explained Variance
FULL TLTM SCALE	.592; 13.21	.823; 17.82		
Conscientiousness	.652; 42.41	.642; 41.91	.652; 42.41 ¹	.642; 41.91 ¹
<i>Planning and structuring</i>	-	-		
1: Following a lesson plan				
<i>Managing classroom behaviour</i>	.627; 73.18	.596; 71.24		
2: Making rules and agreements clear				
3: Intervening when rules are broken				
<i>Managing attention and participation</i>	.557; 69.68	.376; 61.63		
4: Letting the students participate actively				
5: Addressing students who lose attention				
Agreeableness	.337; 25.07	.633; 32.77		
<i>Trusting students</i>	-	-		
6: Trusting that students do not cheat				
<i>Proactive perspective taking</i>	.530; 68.14	.580; 70.67	.586; 54.89 ²	.602; 55.82 ²
7: Paying attention to the needs of students				
8: Supporting students who are struggling mentally				
<i>Connecting and caring</i>	.668; 75.16	.453; 64.67	.668; 75.16 ³	.453; 64.67 ³
9: Building a close bond with students				
10: Being involved in students' personal lives				
<i>Mindset about students' learning</i>	.456; 65.26	.566; 70.10	.456; 65.26 ⁴	.566; 70.10 ⁴
11: Believing that every student can get good grades				
12: Believing that every student can learn				
Extraversion	.510; 44.18	.391; 36.03		
<i>Cheerful engagement</i>	.667; 75.03	.523; 67.75	.667; 75.03 ⁵	.523; 67.75 ⁵
13: Using personal anecdotes in class				
14: Using humour in class				
<i>Communication and presenting</i>	-.086; 52.37	.262; 57.57		
15: Paying attention to the pace of speech				
16: Providing detailed explanation				

Continued on the next page.

Table 3 (Continued)

TLTM item (with item number)	PER DOMAIN AND FACET		FINAL SUBSCALES	
	Teachers	Students	Teachers	Students
	α ; Explained Variance	α ; Explained Variance	α ; Explained Variance	α ; Explained Variance
Openness	.620; 23.38	.732; 29.65	.621; 25.31 ⁶	.711; 30.63 ⁶
<i>Passion for learning and understanding</i>	.485; 67.75	.425; 63.49		
17: Questioning things				
18: Exploring a problem				
<i>Creative teaching</i>	.658; 75.05	.575; 70.20		
19: Experimenting with new teaching methods				
20: Making teaching materials creative and attractive				
<i>Appreciating diversity</i>	.155; 54.60	.349; 60.58		
21: Being open to different opinions				
22: Addressing diversity issues				
<i>Self-reflection and professional development</i>	.369; 61.57	.469; 65.61		
23: Gauging how students experience the lesson				
24: Wanting to improve oneself as a teacher				
<i>Flexibility in teaching</i>	.065; 51.87	.310; 59.19		
25: Adapting when the lesson goes differently than expected				
26: Tailoring the teaching method to each student				
Neuroticism	.397; 36.48	.608; 39.32	.563; 59.52 ⁷	.513; 50.65 ⁷
<i>Stress resilience and recovery</i>	.566; 60.80	.462; 48.39		
27: Starting the lesson optimistically				
28: Standing in front of the class relaxed and at ease				
29: Shaking off stressful situations				
<i>Emotional self-control</i>	.459; 66.92	.423; 64.03		
30: Having patience				
31: Controlling negative emotions				

Note. Domains are displayed in **bold**, facets in *italics*. Each item is a five-point Likert scale ranging from - 2 (too little) to + 2 (too much).

After the factor analyses, seven final subscales were chosen, as shown in the last two columns. Values of $\alpha \geq .600$ are considered acceptable given the developmental stage of the TLTM Scale and the exploratory nature of this research (Taber, 2018). Values of Explained Variance $\geq 50\%$ are considered acceptable for multi-dimensional scale development (Streiner, 1994), but deviations are tolerable when theoretical and internal consistency support retention. For the selected subscales, the item groupings are listed below, followed by the newly assigned subscale names (after the colon). TLTM = Too Little/Too Much; α = Cronbach's alpha.

¹ 'Conscientiousness' items: Classroom organisation and management.

² 'Proactive perspective taking' items plus item 26: Student-centred responsiveness.

³ 'Connecting and caring' items: Connecting and caring.

⁴ 'Mindset about students' learning' items: Mindset about students' learning.

⁵ 'Cheerful engagement' items: Cheerful engagement.

⁶ 'Openness' items without item 26: Reflective and innovative teaching.

⁷ 'Neuroticism' items without items 29 and 30: Emotionally grounded presence.

Two subscales (namely 'Classroom organisation and management' and 'Reflective and innovative teaching') fell below the 50% benchmark. Their retention was justified based on strong conceptual coherence and acceptable internal consistency. Internal consistency exceeded .60 for most subscales in the teacher dataset. In the student dataset, reliability estimates were more mixed, with only some subscales reaching acceptable thresholds.

Among the retained subscales, 'Classroom organisation and management' and 'Reflective and innovative teaching' demonstrated acceptable internal consistency across both groups ($\alpha = .652$ and $.621$ in teachers; $\alpha = .642$ and $.711$ in students). Other subscales performed more variably. For instance, 'Student-centred responsiveness' was slightly below the .60 threshold in the teacher data ($\alpha = .586$) but acceptable in the student data ($\alpha = .602$). In contrast, 'Connecting and caring', 'Mindset about students' learning', 'Cheerful engagement', and the revised 'Emotionally grounded presence' subscale showed lower internal consistency in the student sample (α ranging from $.453$ to $.566$).

Based on these results, several theoretically expected domains were at least partially supported, but others required significant adaptation or could not be reliably confirmed. Full details of reliability coefficients, explained variances, and subscale content are presented in Table 3. Overall, Hypothesis 1 received partial support.

RQ2: Self-Other Agreement in SEMS Assessment

Model selection. Table 4 summarises AIC and BIC values for Models 1 through 3, and adjusted ICC values for Model 2. Model 1 was never the best-fitting model. This indicates that student ratings were meaningfully shaped by the class to which they belonged, both in terms of average levels and in the strength of association with teacher scores. This pattern is theoretically expected, as students evaluate a specific teacher within a shared classroom context.

Model 2 (random intercept only) provided the best fit for the subscales 'Classroom organisation and management', 'Student-centred responsiveness', and 'Cheerful engagement', as indicated by the lowest AIC and BIC values. For 'Connecting and caring', 'Reflective and innovative teaching', and 'Emotionally grounded presence', Model 3 (random intercept and slope) showed superior fit based on AIC, although BIC values were slightly higher.

For the subscale 'Mindset about students' learning', Model 3 had a marginally lower AIC than Model 2, but the improvement was minimal. Following the principle of parsimony, Model 2 was selected for interpretative purposes.

Table 4*Model Fit and ICC Comparison Across TLTM Subscales*

TLTM subscale	Model 1		Model 2			Model 3	
	AIC	BIC	AIC	BIC	ICC	AIC	BIC
Classroom organisation and management	1007.111	1021.224	946.864	965.682	.170	955.192	983.418
Student-centred responsiveness	1204.841	1218.954	1131.789	1150.607	.159	1135.185	1163.412
Connecting and caring	1463.823	1477.936	1382.150	1400.967	.177	1376.246	1404.473
Mindset about students' learning	1211.992	1226.105	1169.678	1188.496	.123	1169.501	1197.728
Cheerful engagement	1691.177	1705.290	1551.511	1570.328	.267	1553.799	1582.026
Reflective and innovative teaching	749.688	763.801	682.738	701.556	.160	679.228	707.455
Emotionally grounded presence	1147.086	1161.199	1096.580	1115.398	.146	1093.618	1121.844

Note. AIC is focused on finding the model that best predicts the data, BIC on selecting the model that is closest to the 'true model' (under the assumption that such a model exists among the set of candidates). In both cases, a lower value indicates a better fit. The lowest value of each fit index between the 3 models is shown in black, the other (higher) values are greyed out. Adjusted ICCs are derived only from Model 2, following the recommendations of Kreft and De Leeuw (1998), Hox (2010), and Hox et al. (2018). TLTM = Too Little/Too Much; AIC = Akaike information criterion; BIC = Bayesian information criterion; ICC = (Adjusted) intraclass correlation coefficient.

Model 1 = Crude model.

Model 2 = Model with random intercept.

Model 3 = Model with random intercept and random slope.

Linear mixed models and correlation analysis. To examine the alignment between secondary school teachers' self-assessed social and emotional skills (SEMS) and their students' evaluations, linear mixed models were estimated using the non-aggregated (long-format) dataset. These models were the same as those introduced in the model selection section (see Table 4) and accounted for the nested structure of the data. The goal was to test whether the associations between teacher and student scores held across individual classrooms.

'Cheerful engagement' consistently showed the highest level of agreement, yielding significant effects across all four models. For other subscales, significance varied depending on the model specification. For example, 'Reflective and innovative teaching' and 'Student-centred responsiveness' showed significant effects in some models but not others, suggesting that classroom context or sample composition may moderate the level of agreement.

To complement these findings, Pearson correlations (r) were also computed using the aggregated (wide-format) dataset. These correlations offer a straightforward indication of overall alignment between teacher and student ratings at the class level (see Table 5 and Appendix C). Correlations ranged from negligible to moderate. The strongest correspondence was again found for 'Cheerful engagement' ($r = .463$), followed by 'Classroom organisation and management' ($r = .294$) and 'Emotionally grounded presence' ($r = .241$). Other subscales showed weaker alignment: 'Student-centred responsiveness' and 'Reflective and innovative teaching' yielded small correlations ($r = .231$ and $r = .220$, respectively), while 'Connecting and caring' ($r = .043$) and 'Mindset about students' learning' ($r = .006$) showed virtually no correspondence.

Table 5*Correlations and Regression Results for the Seven TLTM Subscales*

TLTM subscale	r	Model 1				Model 2				Model 3				Model 4			
		β	p	95% CI		β	p	95% CI		β	p	95% CI		β	p	95% CI	
Classroom organisation and management	.294	.260	<.001	.174	.346	.233	.007	.068	.399	.269	<.001	.138	.399	.131	.083	-.018	.279
Student-centred responsiveness	.231	.113	.010	.027	.198	.129	.099	-.025	.282	.127	.083	-.017	.271	.071	.331	-.074	.215
Connecting and caring	.043	.003	.945	-.084	.090	.025	.755	-.136	.186	.144	.055	-.004	.292	.108	.183	-.065	.281
Mindset about students' learning	.006	.019	.766	-.105	.143	.010	.920	-.184	.204	.004	.983	-.361	.368	-.019	.828	-.189	.152
Cheerful engagement	.463	.397	<.001	.299	.494	.400	<.001	.199	.600	.458	<.001	.290	.626	.349	<.001	.154	.543
Reflective and innovative teaching	.220	.131	.003	.044	.217	.129	.101	-.026	.284	.250	.016	.050	.450	.246	.011	.060	.433
Emotionally grounded presence	.241	.212	<.001	.116	.309	.183	.032	.016	.350	.113	.069	-.009	.234	.082	.714	-.388	.553

Note. r represents the Pearson correlation between the teacher's score (self-evaluation) and the average class score (teacher evaluation). Correlation coefficients should be interpreted as follows: $|r| < .100$ indicates a negligible effect; $.100 \leq |r| < .300$ a weak effect; $.300 \leq |r| < .500$ a moderate effect; $|r| \geq .500$ a strong effect. These thresholds are based on Cohen's recommendations for interpreting observed effect sizes (Cohen, 1988; Cohen, 1992). As for the mixed models, non-significant values ($p > .05$) are greyed out. TLTM = Too Little/Too Much; r = Pearson correlation coefficient; β = Beta coefficient estimate; p = p -value for statistical significance; CI = confidence interval.

Model 1 = Crude model.

Model 2 = Model with random intercept.

Model 3 = Model with random intercept and random slope.

Model 4 = Model with covariates, added to control for teacher- and student-level differences. Teacher-level covariates: gender, tenure, class size, and stage of the class (with the 3rd stage also including the 7th grade). Student-level covariate: gender. Covariates were added to Model 2 for the 'Classroom organisation and management', 'Student-centred responsiveness', 'Mindset about students' learning', and 'Cheerful engagement' subscales; for 'Connecting and caring', 'Reflective and innovative teaching', and 'Emotionally grounded presence', they were added to Model 3.

Conclusion RQ2. Overall, the results reveal limited but variable alignment between teacher and student assessments of social-emotional functioning. Agreement was strongest for observable, affective behaviours such as ‘Cheerful engagement’, while more internal or relational aspects – such as ‘Student-centred responsiveness’ or ‘Connecting and caring’ – showed greater divergence.

RQ3: Relationship between Feedback Attitude and Discrepancy

Descriptive statistics indicated generally positive feedback attitudes among teachers, with scores ranging from 2.20 to 4.80 (*Median* = 3.50) on the 5-point scale, suggesting that most participants were satisfied with the feedback received. An overview of the item-level response distributions is presented in Table 6.

Table 6

Quartile Distributions of Feedback Attitude Scores per Item and Total Scale

Feedback attitude item	Q1	Q2	Q3	M	SD
The feedback I received was an accurate evaluation of my performance	2.75	4.00	4.00	3.33	.87
I believe the feedback received was correct	3.00	4.00	4.00	3.40	.86
The feedback I received helped me learn how I can improve my performance	3.00	3.00	4.00	3.24	1.03
Overall, I was satisfied with the feedback I received	3.00	4.00	4.00	3.60	.86
I felt quite good with the performance feedback I received	3.00	3.00	4.00	3.33	.87
Total feedback attitude scale	2.95	3.50	3.85	3.38	.67

Note. Participants (teachers) responded to translated (Dutch) items on a five-point Likert scale ranging from 1 (Totally disagree) to 5 (Totally agree), with higher scores indicating more favourable feedback attitudes. The ‘total feedback attitude scale’ reflects a global score, calculated as the average of the five items. Q1 = first quartile; Q2 = median; Q3 = third quartile; *M* = mean; *SD* = standard deviation.

Linear regression analysis (see Table 7) was conducted to examine whether discrepancies between teacher and student TLTM scores were linked to teachers’ feedback attitudes. Given the small sample size and the number of predictors, particular attention was paid to Model A, which included only the seven TLTM scores. This model explained a modest proportion of variance (Adjusted $R^2 = .159$).

Two subscales emerged as noteworthy. First, discrepancy in ‘Emotionally grounded presence’ was marginally significant ($\beta = .369$, $p = .052$), indicating that when teachers rated themselves as more emotionally regulated than their students did, they tended to report more positive feedback attitudes. Second, discrepancy in ‘Cheerful engagement’ showed a negative association ($\beta = -.334$, $p = .062$), suggesting that teachers who saw themselves as more cheerful than their students did were somewhat less satisfied with the feedback.

As the study is based on a small sample with many predictors, marginally significant findings ($p < .10$) were interpreted cautiously to avoid overly strict exclusion of potentially meaningful patterns.

Model B, which included additional teacher- and class-level control variables, showed a poorer fit (higher AIC and BIC values), likely due to model complexity. Nevertheless, the two key effects remained: discrepancy in 'Cheerful engagement' reached significance ($p = .047$), and discrepancy in 'Emotionally grounded presence' remained marginally significant ($p = .080$), indicating robustness of these associations across both models.

Table 7

Linear Regression Results Predicting Teacher Feedback Attitude from TLTM Discrepancy Scores

TLTM discrepancy* subscale	Model A (AIC = -34.203; BIC = -20.302)				Model B (AIC = -28.925; BIC = -8.073)			
	β	p	95% CI		β	p	95% CI	
Classroom organisation and management	.145	.361	-.324	.867	.142	.396	-.367	.902
Student-centred responsiveness	-.220	.231	-.912	.228	-.244	.244	-1.030	.273
Connecting and caring	-.131	.413	-.638	.268	-.122	.495	-.682	.337
Mindset about students' learning	.046	.775	-.492	.655	.067	.691	-.490	.731
Cheerful engagement	-.334	.062	-.966	.026	-.387	.047	-1.082	-.007
Reflective and innovative teaching	-.227	.250	-1.247	.335	-.131	.552	-1.152	.627
Emotionally grounded presence	.369	.052	-.005	1.332	.359	.080	-.081	1.374

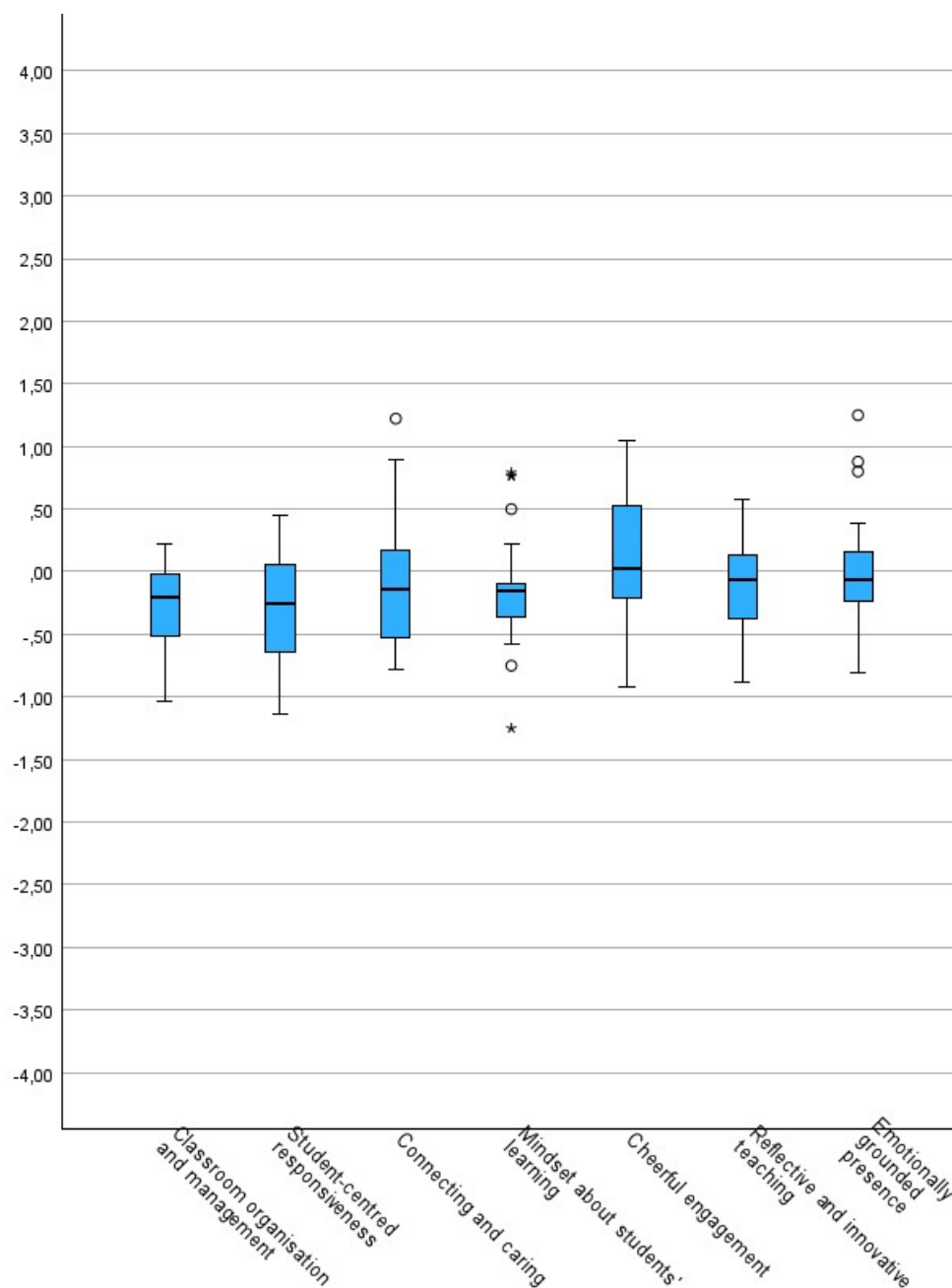
Note. For the linear regression, non-significant values ($p > .10$) are greyed out. TLTM = Too Little/Too Much; AIC = Akaike information criterion; BIC = Bayesian information criterion; β = Beta coefficient estimate (standardised); p = p -value for statistical significance; CI = confidence interval.

Model A = Crude model.

Model B = Model with covariates, added to control for teacher- and student-level differences. Teacher-level covariates: gender, tenure, class size, and stage of the class (with the 3rd stage also including the 7th grade).

*Teacher self-rating minus class average.

To better understand the range and distribution of the discrepancy scores, boxplots were created for each TLTM subscale (see Figure 2).

Figure 2*Boxplots of Discrepancy Scores Across TLTM Subscales*

Note. Discrepancy scores reflect the difference between teacher self-ratings and class average student ratings (teacher minus student). Scores close to 0 indicate alignment; positive values indicate that teachers rated themselves higher than their students did, and negative values indicate the reverse. Although the possible range was -4 to +4, observed discrepancies fell mostly between -1 and +1. TLTM = Too Little/Too Much.

Conclusion RQ3. The analyses provide tentative evidence that discrepancies in specific SEMS domains, particularly 'Emotionally grounded presence' and 'Cheerful

engagement', may influence how feedback is received. However, the small sample size, low magnitude of discrepancies, and heterogeneity across classrooms limit the strength of conclusions. Nonetheless, the overall high satisfaction with feedback and limited discrepancy range point to generally positive feedback experiences, irrespective of perfect alignment.

Discussion

Reflection on Results per Research Question

The results of this study offer preliminary but informative insights into the use of multisource feedback in assessing teachers' social and emotional skills (SEMS). When interpreting these findings, it is important to consider that effect sizes and internal consistency values were evaluated within an exploratory framework. Rather than relying strictly on conventional thresholds (e.g., for Pearson's r or Cronbach's α), findings should be interpreted in relation to the context, the nature of the constructs, and the developmental diversity of the student population.

RQ1: Structural validity of the Too Little/Too Much scale. The first research question explored the structural validity of the TLTM-based student assessment of SEMS, guided by the domains and facets proposed by Scheirlinckx et al. (2023). Although items were adapted from this Big Five-based framework, the novel TLTM response format required an exploratory factor analysis to identify a more data-driven structure.

The resulting seven subscales partially reflected the original theoretical model. For instance, the subscale 'Classroom organisation and management' beautifully aligned with the Conscientiousness domain, while 'Connecting and caring' and 'Mindset about students' learning' mapped onto facets of Agreeableness. Other domains, such as Neuroticism, were only partly retained or needed substantial regrouping, suggesting that the original structure did not fully transfer to the TLTM context.

These discrepancies may reflect a combination of methodological and conceptual factors. On the one hand, students' perspectives may differ from teachers' self-conceptualisations, particularly when assessing abstract or internally oriented skills. On the other hand, the TLTM format introduces behavioural nuance by allowing students to indicate both underperformance and overperformance. An important feature – given that excessive behaviour can also hinder effectiveness (e.g., Pierce & Aguinis, 2013; Busse et al., 2016; Vergauwe et al., 2017) – but one that may influence how they interpret and group behaviours. While several subscales met conventional thresholds for internal consistency and explained variance (e.g., 'Classroom organisation and management', 'Reflective and innovative teaching'), others performed more variably across student and teacher data. In

particular, lower internal consistency in some student-rated subscales (e.g., 'Connecting and caring', 'Cheerful engagement') may reflect developmental variation or heterogeneity in students' classroom experiences.

Taken together, these findings provide partial empirical support for the structure theorised by Scheirlinckx et al. (2023), while also highlighting the need for further refinement of item content and response interpretation. They suggest that while core dimensions of teachers' SEMS are recognisable in student feedback, a one-to-one mapping onto theoretical domains may not be realistic when using a nuanced response format such as TLTM.

RQ2: Self-other agreement in SEMS assessment. The second research question explored the degree of alignment between teacher self-assessments and student evaluations of SEMS. Overall, the patterns of correspondence ranged from minimal to moderate, indicating differences between how teachers perceive themselves and how their students experience them. The strongest alignment was observed for the subscale 'Cheerful engagement' ($r = .463$), followed by 'Classroom organisation and management' ($r = .294$) and 'Emotionally grounded presence' ($r = .241$).

These results are consistent with broader findings in the 360-degree feedback literature, where low-to-moderate agreement between self- and other-ratings is common, particularly for less visible attributes. Prior research highlights that such discrepancies are not merely error, but can reflect meaningful differences in perspective (Fleenor et al., 2010). Mackie (2015) further notes that different raters tend to emphasise different behavioural domains – supervisors (teachers) often focus on external performance, while direct reports (students) are more attuned to interpersonal or relational cues. Intrapersonal traits, being less readily observable, therefore tend to yield lower levels of agreement.

The pattern of results further suggests that students are better able to assess more outwardly observable behaviours such as enthusiasm and classroom engagement, while internal or more nuanced traits appear less consistently recognised. This aligns with existing research indicating that self-other agreement (SOA) tends to be stronger for traits that are visible to the feedback-giver. For instance, Atkins and Wood (2002) included a 'no opportunity to assess' option in their study, underscoring the role of behavioural observability in valid assessment. Similarly, Loignon et al. (2024), through the Leadership Arena-Reputation-Identity (LARI) model, argue that trait observability significantly shapes the predictive validity of multisource ratings.

In this context, Bandura's (1977) theory of observational learning offers a valuable perspective: students are more likely to internalise teachers' social-emotional skills when these are visibly modelled in the classroom. Teachers who openly demonstrate emotional

regulation, empathy, and social engagement are not only more accurately perceived by students, but also play a crucial developmental role.

Notably, even within a single classroom – where students share the same teacher and general context – perceptions of teacher behaviour varied widely. Despite this shared environment, SOA remained limited, suggesting that student evaluations are highly individualised. This unexpected variability reinforces the notion that even outwardly visible behaviours may be interpreted differently depending on personal, relational, or contextual factors.

Given this, the conclusions drawn for RQ2 should be interpreted with caution. Considerable heterogeneity was observed in participants' responses, both across and within classrooms. While the TLTM scale was designed to support nuanced assessments, its format allows for diverse interpretations, which may have contributed to inconsistent patterns and limited the overall strength of associations observed.

RQ3: Relationship between feedback attitude and discrepancy. The third research question investigated whether discrepancies between teacher and student assessments were associated with the teacher's attitude towards the feedback they received. The distribution of discrepancy scores was notably narrow, typically ranging between -1 and +1 across all subscales. While this suggests limited numerical divergence at the group level, the minimal to moderate alignment observed in RQ2 implies that subtle yet meaningful differences in perception were still present.

The TLTM scale was explicitly developed to capture such nuances in behavioural perception, moving beyond the linear assumptions embedded in traditional Likert-type scales (Kaiser & Overfield, 2011; Vergauwe et al., 2017). By allowing respondents to indicate whether a behaviour is shown too little, too much, or just right, the scale acknowledges that behavioural extremes may be dysfunctional (Pierce & Aguinis, 2013) and reduces pressure to choose them. This framing may help explain why discrepancies between teacher and student ratings in this study were often numerically small (e.g., 0 versus 1), and typically situated near the middle ground. These discrepancies nonetheless appear to reflect differing interpretations: rather than indicating categorical disagreement, such slight mismatches may point to diverging judgements about what constitutes 'appropriate' behaviour.

Combined with the generally positive feedback attitudes – no teacher had a mean score below 2.20 – this limited divergence may help explain the absence of stronger or more systematic effects. It also suggests that perfect alignment between student and teacher perceptions may not be required for feedback to be experienced as constructive.

At the same time, the substantial variability in student responses within and between classrooms complicates interpretation and may have obscured meaningful effects at the aggregate level. Nevertheless, two interesting patterns emerged. First, a negative association was found between discrepancy in 'Cheerful engagement' and feedback attitude: teachers who rated themselves as more expressive or cheerful than their students did tended to report less favourable feedback attitudes. This may reflect a disconfirmatory experience, wherein the feedback challenges the teacher's self-perception and reduces receptivity (cf. Devos et al., 2013). Second, teachers who perceived themselves as calmer than their students did (i.e., discrepancy in 'Emotionally grounded presence') reported slightly more positive feedback attitudes. Although marginal, this pattern suggests that certain forms of overestimation – particularly when aligned with positive self-views – may be experienced as reassuring or affirming (cf. Mackie, 2015).

Taken together, these findings underscore that the direction of the discrepancy, rather than its absolute magnitude, may shape how feedback is interpreted and internalised.

Limitations of This Master's Thesis

While this study provides valuable contributions to the literature on multisource feedback and teacher development, several considerations should be kept in mind when interpreting its findings. These relate to the sample, the measurement instruments, and broader methodological and statistical aspects. They offer valuable learning points to guide future research and development efforts.

Sample characteristics and external validity. Although the final dataset included a diverse set of teachers and students, the number of participating teachers was modest. Data cleaning, while necessary to ensure quality, further reduced the number of usable responses. Voluntary participation and an overrepresentation of certain grades and programmes (notably General Education = ASO) may limit the generalisability of the findings. Nevertheless, the variety of school stages and subject areas enhances the relevance of the results to the Flemish secondary education context. It is worth noting that this study aimed to explore emerging patterns within a well-defined sample, rather than to produce population-level generalisations.

Measurement format and interpretability. The TLTM scale, while innovative and pedagogically rich, introduced challenges in terms of interpretability – particularly for younger respondents. As a relatively novel response format, it may have required additional scaffolding or explanation to ensure consistent understanding. Although item misinterpretation was limited, the scale's atypical structure and lack of reverse-coded items reduce its comparability with more traditional psychometric instruments. Some item

groupings showed less conceptual coherence, indicating areas where further refinement could enhance both usability and validity.

Psychometric considerations. From a psychometric perspective, this study relied on Cronbach's α to assess internal consistency. While commonly used (Dunn et al., 2014), this measure is sensitive to scale length and may not fully reflect the conceptual strength of item groupings. In several subscales, α values approached .60 – generally acceptable in exploratory contexts (Taber, 2018) – though likely influenced more by item quantity and response patterns than by high inter-item correlations. Rather than indicating a flaw, these lower α values may reflect the broad and complex nature of the constructs being assessed, suggesting the need for continued refinement of the scale's internal structure.

Multilevel structure and ICC interpretation. The use of hierarchical models was well-suited to the nested data structure, and ICC values (.123 to .267) provided support for multilevel modelling. However, the shared classroom context may have influenced fixed-effect estimates, while within-class variation – largely obscured in aggregated analyses – could have further affected sensitivity. Future research may benefit from more nuanced modelling strategies that capture heterogeneity in classroom-level dynamics and student perspectives more precisely.

Scale design and the bandwidth-fidelity trade-off. The TLTM scale was originally designed to capture a broad conceptual range within each SEMS domain, based on the Big Five-aligned framework proposed by Scheirlinckx et al. (2023). However, the final subscales retained in this study emerged from empirical item clustering and reflect blended groupings that prioritise conceptual breadth over strict construct fidelity. This approach illustrates the classic bandwidth-fidelity dilemma (BDF) (Cronbach & Gleser, 1957; Cronbach, 1990), which posits a trade-off between the breadth (bandwidth) and specificity (fidelity) of a measure. Broader subscales, while potentially offering stronger predictions of general outcomes (Ones & Viswesvaran, 1996), often introduce greater heterogeneity, which may lower internal consistency and obscure fine-grained distinctions. For example, the subscale 'Reflective and innovative teaching' includes items related to curiosity, experimentation, openness to diversity, and self-improvement – originally linked to distinct facets of Openness (to Experience) – and may be considered a compound or blended facet in BFD terms (Salgado, 2017). Although such blending supports broader coverage of teacher behaviour, it complicates efforts to isolate the unique predictive value of narrower facets (Ashton et al., 2014).

Methodological reflections. The quantitative survey method enabled broad participation and systematic data collection, but may not have been the most suitable

approach at this early stage of instrument development. The conceptual tensions observed in RQ1 persisted through the analyses of RQ2 and RQ3, suggesting that a return to scale construction may be warranted. A more inductive, qualitative approach – such as involving students and teachers in item generation through interviews or focus groups – could strengthen the validity and interpretability of the instrument.

The use of discrepancy scores (RQ3) also warrants reflection. Given that 'scoring higher' may reflect either under- or overestimation – and that the meaning of 'too much' and 'too little' may vary by respondent – such scores may not capture the full nuance of student-teacher misalignment. Identical discrepancy values can arise from very different combinations of student and teacher responses, which may lead to distinct interpretations and reactions. Moreover, teachers received percentage breakdowns of student responses across TLTM categories ('too little', 'just right', 'too much'), rather than raw scores. This feedback design may have shaped feedback attitudes more through general impressions than precise comparisons, further limiting the interpretive value of discrepancy scores.

Strengths and Practical Implications

Despite the exploratory nature of the findings, this study offers several noteworthy strengths and points of practical relevance for the educational field.

A new lens on SEMS assessment. First, the longitudinal, multisource design allowed for a unique comparison between teacher self-assessments and student ratings of SEMS across time. By incorporating student perspectives – using the innovative TLTM format – this study contributes to a growing body of international research advocating for more participatory approaches in teacher feedback and professional development. Relevant studies have emerged from diverse contexts, including the Philippines, the United States, Spain, and Turkey (Morales, 2016; Darling-Hammond et al., 2017; García, 2024; Alkan et al., 2024). Closer to home, research from Belgium and the Netherlands underscores similar themes, with studies highlighting the need for more collaborative, feedback-rich teacher development (e.g., van der Lans et al., 2018; Tuytens et al., 2019; Minea-Pic et al., 2021).

Understanding feedback attitudes. Second, the detailed analysis of SOA (RQ2) and its relation to teachers' feedback attitudes (RQ3) addresses a topic that remains underexplored in educational literature. While the observed effects were modest, they nonetheless revealed subtle but meaningful dynamics: teachers who perceived themselves as more emotionally regulated than their students did were somewhat more receptive to feedback, whereas perceived overexpressiveness was linked to reduced satisfaction. Importantly, this study offers a rare lens into the emotional and attitudinal responses to perceptual misalignment in educational contexts. As such, it adds nuance to the broader

feedback orientation literature (e.g., London & Smither, 2002; Linderbaum & Levy, 2010; Yang et al., 2023), underscoring that perceptual congruence may shape not only performance-related outcomes but also teachers' receptivity to feedback itself. These findings may inform strategies for facilitating reflective, emotionally attuned feedback dialogues between students and teachers, and align with broader research on the role of emotional regulation, personality, and well-being in teaching (e.g., Wang et al., 2023).

Lessons for Too Little/Too Much implementation. Finally, the study raises important methodological considerations for future research and practice. While the TLTM format shows promise as a pedagogically rich feedback tool, its novel structure may require more explicit explanation, particularly for younger respondents. Complementary approaches – such as guided reflection or open-ended comments – could strengthen the feedback process by adding qualitative depth. This study thus encourages a more nuanced, student-inclusive approach to teacher development – one that recognises both the potential and the limitations of multisource feedback, and lays the groundwork for future educational research and evidence-based practice.

Further Research

Given the exploratory nature of this study, several avenues for future research emerge, particularly around the development of the TLTM scale and the conditions under which student feedback contributes meaningfully to teacher development.

Scale development and validation. A clear priority is the further refinement of the TLTM scale. Cognitive interviewing (e.g., Hsu et al., 2025) and think-aloud protocols (e.g., Trenor et al., 2011) could offer valuable insights into how students interpret TLTM items, particularly in light of age-related variation in comprehension. Think-aloud protocols may themselves be considered a form of cognitive interviewing (Willis, 1994, as cited in McColl, 2006; Miller et al., 2014, as cited in Arthur et al., 2021), further supporting their use in this context. Item rewording, clearer anchoring of scale points, and restructuring TLTM subscales for stronger internal coherence may improve the scale's psychometric properties.

Analytical approaches to (mis)alignment. While discrepancy scores served as a useful first step in operationalising student-teacher (mis)alignment, future research may benefit from more refined metrics that account for the direction, valence, or distribution of ratings across the TLTM categories. For instance, future studies could distinguish between types of divergence (e.g., when teachers rate themselves as 'just right' and students as 'too little' versus the reverse), as these may elicit different reactions and implications for feedback acceptance.

Moderator variables. This study included a small number of exploratory models with control variables, but moderators were not systematically investigated. Future research should integrate moderators more explicitly into the study design. Variables such as teacher personality traits, experience, class size, student engagement, or perceived credibility of the feedback provider may shape both SOA and the impact of feedback. Moderated models could clarify when and for whom student feedback is most meaningful.

Longitudinal and comparative designs. While this study collected data across time points, only short-term effects were examined. Longitudinal research tracking the impact of student feedback on actual teacher behaviour, classroom dynamics, or student learning outcomes remains a crucial next step. Comparative studies across schools or countries may also help identify which contextual factors promote or hinder effective use of student-based feedback.

In sum, this thesis lays the groundwork for a more evidence-based, inclusive feedback culture in education. By refining measurement tools, investigating key moderators, and deepening our understanding of both student and teacher perspectives, future research can enhance the design and impact of student-informed feedback systems.

Additional Reflections: Factors in the Relationship between Student Feedback and Teacher Improvement

The findings of this study highlight the complexity of implementing student feedback as a meaningful tool for teacher development. While the core analyses focused on structural validity, SOA, and feedback attitude, the literature reviewed and additional observations suggest that several factors may condition the effectiveness of student feedback in practice.

Feedback implementation. The effectiveness of feedback depends not only on its content but also on its timing and format. As highlighted in the literature, feedback that is timely and aligned with specific achievement goals can significantly enhance performance (Anseel et al., 2010). In the present study, feedback reports were returned as promptly as possible. However, they constituted a relatively isolated event. More systematic integration of student feedback into formative assessment cycles could promote sustained reflection and professional growth.

The format of feedback is equally consequential. While Likert-type items are widely used, they are susceptible to common rating biases, such as halo effects and leniency (Brown et al., 2017). In contrast, open-ended comments have been shown to yield richer, more specific, and actionable insights, making them particularly valuable for formative purposes (Omer et al., 2023). The TLTM format, employed in this study, represents an

innovative attempt to mitigate some of the limitations of traditional rating scales by prompting respondents to indicate when a behaviour is shown ‘too little’, ‘too much’, or ‘just right’. This approach may help counteract rating inflation and encourage more deliberate evaluations. Nonetheless, its unfamiliar structure and pedagogical framing may require additional scaffolding to ensure interpretability, particularly among younger respondents. Forced-choice rankings have also been proposed as an alternative, as they require more nuanced differentiation between behavioural traits (Brown et al., 2017). Future research should investigate how different feedback formats (e.g., TLTM vs. Likert or open-ended) and delivery modes (e.g., written reports vs. interactive coaching) influence teachers’ engagement with feedback and its perceived usefulness.

Feedback culture and goal-setting. Beyond implementation, institutional conditions and mechanisms play a critical role in shaping feedback uptake. A supportive feedback culture – one that aligns feedback with specific achievement goals and fosters openness to reflection – can positively influence how feedback is received and acted upon (London & Smither, 2002; Yang et al., 2023). Goal-setting is often posited as a mechanism that may help link feedback to behavioural change. In educational research, for instance, Criss et al. (2023) included goal-setting as part of a feedback intervention aimed at improving classroom management. However, the study could not isolate the effects of this component. Whether such strategies are effective with more abstract domains like SEMS – particularly when the feedback originates from students – remains an open question.

Feedback follow-up. Feedback must be processed, discussed, and translated into action. Research suggests that active engagement with feedback is associated with significant gains in achievement (Brooks et al., 2021). Structured follow-up mechanisms, such as those described by Husain and Khan (2016), can further enhance the utility of student evaluations by supporting this engagement. In this study, while teachers were invited to set developmental goals, the extent of follow-up support was not systematically assessed. Institutions play a central role in this process by offering resources, cultivating a culture of reflective practice, and embedding feedback within broader professional growth trajectories (Karm et al., 2022; Cornes et al., 2023). Future studies could explore the added value of follow-up structures such as peer discussion and institutional framing in shaping feedback uptake and perceived usefulness.

Student perceptions and anonymity. The perceived relevance and authenticity of student feedback may be influenced by multiple factors, including students’ familiarity with the teacher’s role (Aguinis et al., 2012) and their experience of the feedback process itself. In classroom settings, students typically observe their teachers’ behaviours on a daily basis and

are thus well-positioned to provide meaningful input. However, power dynamics may complicate this role. Anonymity can mitigate fears of repercussions and increase honesty, especially in hierarchical settings (Caulfield, 2007). In this study, TLTM ratings were collected anonymously, yet no qualitative data were gathered to examine whether students actually felt safe or motivated to respond authentically. Future research could use interviews or open-ended survey items to explore how students experience their role as feedback-givers, what supports or undermines their willingness to participate, and how classroom feedback cultures shape the perceived usefulness of such processes.

Conclusion

This master's thesis examined the potential of student-informed assessments in evaluating teachers' social and emotional skills (SEMS), complementing traditional self-reports with the innovative Too Little/Too Much (TLTM) scale. Through a longitudinal, multisource design, the study addressed three central research questions: the structural validity of the TLTM scale (RQ1), the degree of self-other agreement (SOA) between teachers and students (RQ2), and the relationship between perceptual (mis)alignment and teachers' feedback attitudes (RQ3).

Findings offer partial but meaningful support for the TLTM scale's application in educational contexts. While some subscales showed acceptable reliability and theoretical alignment, others displayed inconsistencies, particularly in student ratings. Observable behaviours – such as humour and the use of personal anecdotes – were assessed more consistently than internal or relational competencies, highlighting the challenges of capturing nuanced teacher behaviours through student surveys.

SOA was generally modest, with the highest agreement found for visible behaviours. These results suggest that while students can offer valuable insights into certain aspects of teacher functioning, such feedback must be interpreted with contextual and developmental sensitivity, particularly in formative applications.

Analyses of discrepancy and feedback attitudes revealed subtle but thought-provoking patterns. Teachers who saw themselves as more expressive than students did tended to respond less favourably to feedback, whereas those who rated themselves as calmer than students did showed slightly more positive attitudes. Nevertheless, teachers generally responded constructively, even in the presence of minor discrepancies. This suggests that complete alignment is not a prerequisite for meaningful engagement.

Taken together, these findings support the integration of student perspectives in teacher development, provided that feedback mechanisms are carefully designed,

pedagogically scaffolded, and supported by an enabling institutional environment. The study thus lays a promising foundation for further research and practice.

Future work should focus on refining the TLTM scale, identifying relevant moderator variables (e.g., personality, feedback format), and adopting longitudinal or mixed-methods approaches to better understand the dynamics of feedback reception and use. By advancing these lines of inquiry, both scholars and practitioners can contribute to the development of more inclusive, evidence-informed feedback systems that support teacher growth and foster stronger teacher-student relationships.

In sum, this thesis advocates for a more participatory and reflective approach to teacher development – one that recognises students not only as learners, but as meaningful contributors to the educational process.

Statement Regarding the Use of Generative AI

Generative AI (GAI; in this case: the free version of ChatGPT) was used in several phases of this research project. Taking into account the faculty guidelines for the use of AI in the master's thesis, this was done in a responsible manner.

The sources cited in this thesis were not identified or selected through GAI. Instead, Ghent University's access to Web of Science was gratefully used for the literature review, supplemented by occasional searches via Google Scholar.

GAI primarily supported brainstorming during the early stages of the research process, especially when refining research questions and exploring statistical analysis options. All suggestions were critically reviewed and adapted to ensure scientific integrity. GAI was also consulted to help clarify certain statistical concepts.

In addition, the 'rephrase' function was used to improve the readability of certain sections. For translations from Dutch to English, Google Translate was preferred, and GAI was not used for this purpose.

References

- Aesaert, K., Voogt, J., Kuiper, E., & van Braak, J. (2017). Accuracy and bias of ICT self-efficacy: An empirical study into students' over- and underestimation of their ICT competences. *Computers in Human Behavior*, 75(1), 92–102.
<https://doi.org/10.1016/j.chb.2017.05.010>
- Aguinis, H., Gottfredson, R. K., & Joo, H. (2012). Delivering effective performance feedback: The strengths-based approach. *Business Horizons*, 55(2), 105–111.
<https://doi.org/10.1016/j.bushor.2011.10.004>
- Akram, M., & Zepeda, S. J. (2015). Development and validation of a teacher self-assessment instrument. *Journal of Research and Reflections in Education*, 9(2), 134–148.
- Alkan, Ö. K., Aksoy, N. C., Kulaksız, T., Kaplan, H. A., Durmaz, B. N., Özcan, M., & Kalkavan, B. (2024). A multi-feedback system integrated simulation-based teacher training to scaffold pre-service teachers' teaching skills: A phenomenological approach. *Education and Information Technologies*, 29(15), 20691–20713. <https://doi.org/10.1007/s10639-024-12657-4>
- Allen, K., Kern, M. L., Vella-Brodrick, D., Hattie, J., & Waters, L. (2018). What schools need to know about fostering school belonging: A meta-analysis. *Educational Psychology Review*, 30(1), 1–34. <https://doi.org/10.1007/s10648-016-9389-8>
- Anseel, F., Van Yperen, N. W., Janssen, O., & Duyck, W. (2010). Feedback type as a moderator of the relationship between achievement goals and feedback reactions. *Journal of Occupational and Organizational Psychology*, 84(4), 703–722.
<https://doi.org/10.1348/096317910X516372>
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5), 502–508. <https://doi.org/10.1111/opo.12131>

- Arthur, E. K., Menon, U., Browning, K., Overcash, J., & Wills, C. E. (2021). Challenges of cognitive interviewing in sensitive health topic research. *Nursing Research*, 70(5), 376–382. <https://doi.org/10.1097/nnr.0000000000000530>
- Ashton, M. C., Paunonen, S. V., & Lee, K. (2014). On the validity of narrow and broad personality traits: A response to Salgado, Moscoso, and Berges (2013). *Personality and Individual Differences*, 56, 24–28. <https://doi.org/10.1016/j.paid.2013.08.019>
- Atkins, P. W. B., & Wood, R. E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, 55(4), 871–904. <https://doi.org/10.1111/j.1744-6570.2002.tb00133.x>
- Atwater, L. E., & Brett, J. F. (2006). 360-degree feedback to leaders: Does it relate to changes in employee attitudes? *Group and Organization Management*, 31(5), 578–600. <https://doi.org/10.1177/1059601106286887>
- Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45(1), 141–164. <https://doi.org/10.1111/j.1744-6570.1992.tb00848.x>
- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1997). Self-other agreement: Does it really matter? *Personnel Psychology*, 51(3), 577–598. <https://doi.org/10.1111/j.1744-6570.1998.tb00252.x>
- Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Teachers' judgment of students' potential as a function of teachers' susceptibility to biasing information. *Journal of Personality and Social Psychology*, 42(3), 541–547. <https://doi.org/10.1037/0022-3514.42.3.541>
- Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), 411–421.

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bergner, S., Davda, A., Culpin, V., & Rybníček, R. (2016). Who overrates, who underrates? Personality and its link to self-other agreement of leadership effectiveness. *Journal of Leadership & Organizational Studies*, 23(3), 335–354.
<https://doi.org/10.1177/1548051815621256>
- Berson, Y., & Sosik, J. J. (2007). The relationship between self-other rating agreement and influence tactics and organizational processes. *Group & Organization Management*, 32(6), 675–698. <https://doi.org/10.1177/1059601106288068>
- Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*, 310(6973), 170. <https://doi.org/10.1136/bmj.310.6973.170>
- Borg, S., & Edmett, A. (2019). Developing a self-assessment tool for English language teachers. *Language Teaching Research*, 23, 655–679.
<https://doi.org/10.1177/1362168817752543>
- Bramley, G., Campbell-Pilling, K., & Simmons, C. (2020). 'Don't feedback in anger': Enhancing student experience of feedback. *6th International Conference on Higher Education Advances (HEAd'20)*, Universitat Politècnica de València, 1147-1154.
<https://doi.org/10.4995/head20.2020.11216>
- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, J., & Neill, S. (2021). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education*, 105.
<https://doi.org/10.1016/j.tate.2021.103387>
- Brown, A., Inceoglu, I., & Lin, Y. (2017). Preventing rater biases in 360-degree feedback by forcing choice. *Organizational Research Methods*, 20 (1), 121–148.
<https://doi.org/10.1177/1094428116668036>

- Busse, C., Mahlendorf, M. D., & Bode, C. (2016). The ABC for studying the too-much-of-a-good-thing effect: A competitive mediation framework linking antecedents, benefits, and costs. *Organizational Research Methods*, 19(1), 131–153.
<https://doi.org/10.1177/1094428115579699>
- Caulfield, J. (2007). What motivates students to provide feedback to teachers about teaching and learning? An expectancy theory perspective. *International Journal for the Scholarship of Teaching & Learning*, 1(1). <https://doi.org/10.20429/ijsoitl.2007.010107>
- Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and recent advances in model selection. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Handbook of the Philosophy of Science* (pp. 583–605). <https://doi.org/10.1016/b978-0-444-51862-0.50018-6>
- Cieciuch, J., & Strus, W. (2021). Toward a model of personality competencies underlying social and emotional skills: Insight from the circumplex of personality metatraits. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.711323>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
<https://doi.org/10.1037/0033-2909.112.1.155>
- Cornes, S., Torre, D., Fulton, T. B., Oza, S., Teherani, A., & Chen, H. C. (2023). When students' words hurt: 12 tips for helping faculty receive and respond constructively to student evaluations of teaching. *Medical Education Online*, 28(1).
<https://doi.org/10.1080/10872981.2022.2154768>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>

- Criss, C. J., Konrad, M., Alber-Morgan, S. R., Brock, M. E., & Harris, A. B. (2023). Using performance feedback with and without goal setting on teachers' classroom management skills. *Behavioral Disorders*, 49(2), 91–105. <https://doi.org/10.1177/01987429231201096>
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). McGraw-Hill.
- Cronbach, L. J., & Gleser, C. G. (1957). *Psychological tests and personnel decisions*. University of Illinois Press.
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute. <https://doi.org/10.54300/122.311>
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 416–436). Sage Publications. <https://doi.org/10.4135/9781446249215.n21>
- Devos, G., Hulpia, H., Tuytens, M., & Sinnaeve, I. (2013). Self-other agreement as an alternative perspective of school leadership analysis: An exploratory study. *School Effectiveness and School Improvement*, 24(3), 296–315. <https://doi.org/10.1080/09243453.2012.693103>
- Diaries, J. D. (2024, November 15). Choosing the best model: A friendly guide to AIC and BIC. *Medium*. <https://medium.com/@jshaik2452/choosing-the-best-model-a-friendly-guide-to-aic-and-bic-af220b33255f>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>

- Ernst, H. M., Wittwer, J., & Voss, T. (2023). Do they know what they know? Accuracy in teacher candidates' self-assessments and its influencing factors. *British Educational Research Journal*, 49, 649-673. <https://doi.org/10.1002/berj.3860>
- European Commission (2022). *Country reports - Belgium*. Education and Training Monitor 2022. <https://op.europa.eu/webpub/eac/education-and-training-monitor-2022/en/country-reports/belgium.html>
- Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine*, 80(Suppl.), S46–S54. <https://doi.org/10.1097/00001888-200510001-00015>
- Ferguson, D. L., Hanreddy, A., & Draxton, S. (2011). Giving students voice as a strategy for improving teacher practice. *London Review of Education*, 9(1), 55-70. <https://doi.org/10.1080/14748460.2011.550435>
- Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly*, 21(6), 1005–1034. <https://doi.org/10.1016/j.leaqua.2010.10.006>
- Gan, Z. (2020). How learning motivation influences feedback experience and preference in Chinese university EFL students. *Frontiers in Psychology*, 11, 1-14. <https://doi.org/10.3389/fpsyg.2020.00496>
- García, E. C. (2024). Peer feedback for teaching professional development: Conditions for it to take effect. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186x.2024.2391577>
- Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference* (5th ed.). CRC Press.
- Guo, Y., Zhang, Y., Liao, J., Guo, X., Liu, J., Xue, X., Li, C., Zhang, M., & Zhang, Y. (2017). Negative feedback and employee job performance: Moderating role of the big five. *Social Behavior and Personality*, 45(10), 1735–1744. <https://doi.org/10.2224/sbp.6478>

- Ham, S.-H., Duyar, I., & Gumus, S. (2015). Agreement of self-other perceptions matters: Analyzing the effectiveness of principal leadership through multi-source assessment. *Australian Journal of Education*, 59(3), 225–246.
<https://doi.org/10.1177/0004944115603373>
- Hammer, R., Peer, E., & Babad, E. (2018). Faculty attitudes about student evaluations and their relations to self-image as teacher. *Social Psychology of Education*, 21(3), 517–537.
<https://doi.org/10.1007/s11218-018-9426-1>
- Harris, R. J. (1985). *A primer of multivariate statistics* (2nd ed.). Academic Press.
- Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge. <https://doi.org/10.4324/9780203887332>
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780203852279>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.
- Hsu, H., Bautista, G., & Yang, X. (2025). Reassessing the faculty encouragement scale through a cognitive interview approach with first-year engineering students. *European Journal of Engineering Education*, 1–17. <https://doi.org/10.1080/03043797.2025.2475340>
- Husain, M., & Khan, S. (2016). Students' feedback: An effective tool in teachers' evaluation system. *International Journal of Applied and Basic Medical Research*, 6(3), 178-181.
<https://doi.org/10.4103/2229-516x.186969>
- Kaiser, R. B., & Overfield, D. V. (2011). Strengths, strengths overused, and lopsided leadership. *Consulting Psychology Journal: Practice and Research*, 63(2), 89–109.
<https://doi.org/10.1037/a0024470>

- Karm, M., Sarv, A., & Groccia, J. (2022). The relationship between students' evaluations of teaching and academics professional development. *Journal of Further and Higher Education*, 46(8), 1161–1174. <https://doi.org/10.1080/0309877X.2022.2057214>
- Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology*, 85(5), 708–723. <https://doi.org/10.1037/0021-9010.85.5.708>
- Kinicki, A. J., Prussia, G. E., Wu, B. (J.), & McKee-Ryan, F. M. (2004). A covariance structure analysis of employees' response to performance feedback. *Journal of Applied Psychology*, 89(6), 1057–1069. <https://doi.org/10.1037/0021-9010.89.6.1057>
- Korsgaard, M. A. (1996). The impact of self-appraisals on reactions to feedback from others: The role of self-enhancement and self-consistency concerns. *Journal of Organizational Behavior*, 17(4), 301–311. [https://doi.org/10.1002/\(SICI\)1099-1379\(199607\)17:4<301::AID-JOB749>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1099-1379(199607)17:4<301::AID-JOB749>3.0.CO;2-N)
- Kreft, I. G. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. SAGE Publications.
- Lee, A., & Carpenter, N. C. (2018). Seeing eye to eye: A meta-analysis of self-other agreement of leadership. *The Leadership Quarterly*, 29(2), 253–275. <https://doi.org/10.1016/j.leaqua.2017.06.002>
- Li, A., Aidossova, Z., Tazhina, G., & Tatyeva, Z. (2021). Emotional intelligence and self-assessment of school teachers. *Education in a changing world: global challenges and national priorities*, 114, 86-93. <https://doi.org/10.15405/epsbs.2021.07.02.11>
- Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the feedback orientation scale (FOS). *Journal of Management*, 36(6), 1372–1405. <https://doi.org/10.1177/0149206310373145>

- Loignon, A. C., Fleenor, J. W., Jeong, S., & Woehr, D. J. (2024). Does what others can(not) see really matter? The relationship between leadership Arena-Reputation-Identity (LARI) model and leader effectiveness. *Journal of Applied Psychology*, 110(3), 404-431.
<https://doi.org/10.1037/apl0001238>
- London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review*, 12(1), 81–100. [https://doi.org/10.1016/s1053-4822\(01\)00043-2](https://doi.org/10.1016/s1053-4822(01)00043-2)
- Machine Learning Quick Reads (2024, April 7). Model Selection with AIC & BIC. *Medium*. <https://machinelearningabc.medium.com/model-selection-with-aic-bic-10ac9dac4c5a>
- Mackie, D. (2015). Who sees change after leadership coaching? An analysis of impact by rater level and self-other alignment on multi-source feedback. *International Coaching Psychology Review*, 10(2), 118-130. <https://doi.org/10.53841/bpsicpr.2015.10.2.118>
- McColl, E. (2006). Cognitive Interviewing: A tool for improving questionnaire design. *Quality of Life Research*, 15(3), 571–573. <https://doi.org/10.1007/s11136-005-5263-8>
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81–90.
<https://doi.org/10.1037/0022-3514.52.1.81>
- McNaughton, S., Rosedale, N., Zhu, T., Siryj, J., Oldehaver, J., Teng, S. L., Williamson, R., & Jesson, R. (2022). Relationships between self-regulation, social skills and writing achievement in digital schools. *Reading and Writing*, 35, 1201-1219.
<https://doi.org/10.1007/s11145-021-10232-8>
- Menard, S. (1995). *Applied logistic regression analysis* (Vol. 106). SAGE Publications.

- Minea-Pic, A., Nusche, D., Sinnema, C., & Stoll, L. (2021). Teachers' professional learning study: Diagnostic report for the Flemish community of Belgium (*OECD Education Policy Perspectives*, No. 31). OECD Publishing. <https://doi.org/10.1787/7a6d6736-en>
- Miyamoto, K., Huerta, M. C., & Kubacka, K. (2015). Fostering social and emotional skills for well-being and social progress. *European Journal of Education*, 50(2), 147–159. <https://doi.org/10.1111/ejed.12118>
- Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, 16(2), 82–99. <https://doi.org/10.1016/j.jslw.2007.04.002>
- Moralejo, L., Andersen, E., Hilsmann, N., & Kennedy, L. (2019). Measuring student responses in and instructors' perceptions of student evaluation teaching (SETs), pre and post intervention. *The Canadian Journal for the Scholarship of Teaching and Learning*, 10(3). <https://doi.org/10.5206/cjsotl-rcacea.2019.3.8052>
- Morales, M. P. E. (2016). Participatory Action Research (PAR) cum Action Research (AR) in teacher professional development: A literature review. *International Journal of Research in Education and Science*, 2(1), 156-165.
- Murray, H. G. (1997). Does evaluation of teaching lead to improvement of teaching? *International Journal for Academic Development*, 2(1), 8–23. <https://doi.org/10.1080/1360144970020102>
- Napolitano, C. M., Sewell, M. N., Yoon, H. J., Soto, C. J., & Roberts, B. W. (2021). Social, emotional, and behavioral skills: An integrative model of the skills associated with success during adolescence and across the life span. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.679561>

- Nielsen, K., Tafvelin, S., von Thiele Schwarz, U., & Hasson, H. (2021). In the eye of the beholder: How self-other agreements influence leadership training outcomes as perceived by leaders and their followers. *Journal of Business and Psychology*, 37, 73-90.
<https://doi.org/10.1007/s10869-020-09730-3>
- OECD (2015). *Skills for social progress: The power of social and emotional skills*. OECD Publishing. <https://doi.org/10.1787/9789264226159-en>
- Omer, K., Jacobs, S., Cottenie, K., Bettger, B., Dawson, J., Graether, S., Murrant, C., Zettel, J., & Newton, G. (2023). Evaluating and improving the formative use of student evaluations of teaching. *The Canadian Journal for the Scholarship of Teaching and Learning*, 14(1). <https://doi.org/10.5206/cjsotlrcacea.2023.1.10960>
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17(6), 609–626.
- Palk, K. (2018). *Middle school teachers attitudes towards the need for student feedback and attitudes towards the impact of feedback on teachers professional development* (Master's thesis). University of Tartu, 3385-3394. Retrieved from <https://dspace.ut.ee/handle/10062/63303>
- Pichurin, V., Umerenko, V., & Dutko, T. (2023). The formation of emotional stability of students in physical education. *Physical Education and Sports International Conference*, 705-712. <https://doi.org/10.51582/interconf.19-20.04.2023.075>
- Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management*, 39(2), 313-338. <https://doi.org/10.1177/0149206311410060>
- Primi, R., Santos, D., John, O. P., & De Fruyt, F. (2021). *SENNA: Inventory for the assessment of social and emotional skills: Technical manual*. Instituto Ayrton Senna. <https://doi.org/10.31234/osf.io/byvpr>

- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23(2), 146–159.
<https://doi.org/10.1016/j.tate.2006.04.035>
- Salgado, J. F. (2017). Bandwidth-fidelity dilemma. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–4). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1280-1
- Scheirlinckx, J., Van Raemdonck, L., Abrahams, L., Teixeira, K. C., Alves, G., Primi, R., John, O. P., & De Fruyt, F. (2023). Social-emotional skills of teachers: Mapping the content space and defining taxonomy requirements. *Frontiers in Education*, 8.
<https://doi.org/10.3389/feduc.2023.1094888>
- Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, 143, 565–600.
<https://doi.org/10.1037/bul0000098>
- Sosik, J. J., & Godshalk, V. M. (2004). Self-other rating agreement in mentoring: Meeting protégé expectations for development and career advancement. *Group and Organization Management*, 29(4), 442–469. <https://doi.org/10.1177/1059601103257421>
- Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI. *Journal of Personality and Social Psychology*, 123(1), 192–222.
<https://doi.org/10.1037/pspp0000401>
- Stalmeijer, R. E., Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., Peters, W. G., van Coppenolle, L., & Scherpbier, A. J. J. A. (2009). Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Advances in Health Sciences Education*, 15(3), 315–328. <https://doi.org/10.1007/s10459-009-9199-6>

- Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Lawrence Erlbaum.
- Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *The Canadian Journal of Psychiatry*, 39(3), 135–140.
<https://doi.org/10.1177/070674379403900303>
- Stronge, J. H. (2018). *Qualities of effective teachers* (3rd ed.). Association for Supervision & Curriculum Development.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296.
<https://doi.org/10.1007/s11165-016-9602-2>
- Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, 87(2), 320–332. <https://doi.org/10.1037/0021-9010.87.2.320>
- Trenor, J. M., Miller, M. K., & Gipson, K. G. (2011, June). Utilization of a think-aloud protocol to cognitively validate a survey instrument identifying social capital resources of engineering undergraduates. *Paper presented at the 2011 ASEE Annual Conference & Exposition*, Vancouver, BC, Canada.
- Tsui, A. S., & Ashford, S. J. (1994). Adaptive self-regulation: A process view of managerial effectiveness. *Journal of Management*, 20(1), 93–121.
<https://doi.org/10.1177/014920639402000105>
- Tuytens, M., Moolenaar, N., Daly, A., & Devos, G. (2019). Teachers' informal feedback seeking towards the school leadership team: A social network analysis in secondary schools. *Research Papers in Education*, 34(4), 405–424.
<https://doi.org/10.1080/02671522.2018.1452961>

- van der Lans, R., van de Grift, W., & van Veen, K. (2018). Developing an instrument for teacher feedback: Using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *Journal of experimental education*, 86(2), 247-264. <https://doi.org/10.1080/00220973.2016.1268086>
- VanVoorhis, C. R. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43–50. <https://doi.org/10.20982/tqmp.03.2.p043>
- Vergauwe, J., Wille, B., Hofmans, J., Kaiser, R. B., & De Fruyt, F. (2017). The too little/too much scale: A new rating format for detecting curvilinear effects. *Organizational Research Methods*, 20(3), 518-544. <https://doi.org/10.1177/1094428117706534>
- Wang, H., Burić, I., Chang, M., & Gross, J. J. (2023). Teachers' emotion regulation and related environmental, personal, instructional, and well-being factors: A meta-analysis. *Social Psychology of Education*, 26(6), 1651–1696. <https://doi.org/10.1007/s11218-023-09810-1>
- Whitehead, O. (2024, September 1). Belgium's teacher shortage: How bad is it? *The Brussels Times*. <https://www.brusselstimes.com/1200898/belgiums-teacher-shortage-how-bad-is-it>
- Yammarino, F. J., & Atwater, L. E. (1997). Do managers see themselves as other see them? Implications of self-other rating agreement for human resources management. *Organizational Dynamics*, 25(4), 35–44. [https://doi.org/10.1016/s0090-2616\(97\)90035-8](https://doi.org/10.1016/s0090-2616(97)90035-8)
- Yang, L., Wu, Y., Liang, Y., & Yang, M. (2023). Unpacking the complexities of emotional responses to external feedback, internal feedback orientation and emotion regulation in higher education: A qualitative exploration. *Systems*, 11(6), 315-334. <https://doi.org/10.3390/systems11060315>

- Yu, S. O. (2016). Using students' feedback to evaluate teachers' effectiveness. *Journal for Educators, Teachers and Trainers*, 2(1), 156–165. Retrieved from <https://jett.labosfor.com/index.php/jett/article/view/222>
- Zong, Z., Schunn, C., & Wang, Y. (2021). What makes students contribute more peer feedback? The role of within-course experience with peer feedback. *Assessment & Evaluation in Higher Education*, 1–12. <https://doi.org/10.1080/02602938.2021.1968792>

Appendix A

Feedback Report Examples

This appendix presents some screenshots from a feedback report, made in Dutch, to give an idea of what such a report looks like. In addition to an overview of current skills (see Figure A1), they also contain development opportunities per Big Five domain, under which teachers' social and emotional skills (SEMS) are categorised in the framework of Scheirlinckx et al. (2023) (see Figure A2).

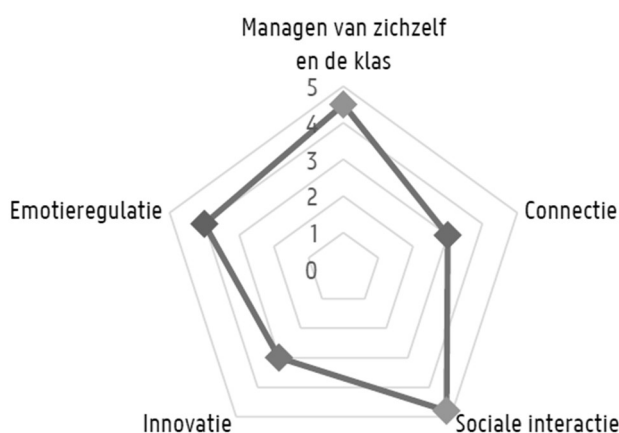
Note that the 'current skills' section is based on the 'SEMS complete' questionnaire, not on the 'SEMS short – TLTM' (self-rating) scale.

As for the development opportunities, two facets are presented for each domain, even when the domain consists of more than two facets. The selection of these facets is deliberately non-analytical, leaning instead on practical and pedagogical considerations to enhance the utility and clarity of the reports. Two main criteria are used:

1. Visibility and assessability: The selected facets are noticeable and assessable by students, guaranteeing that the feedback is based on clearly observable teacher actions.
2. Differentiation of content: The facets are chosen to offer a wide range of informative content, allowing teachers to gain varied insights that support thorough professional development. This means that (where possible) no two facets are chosen that are close in meaning.

Figure A1

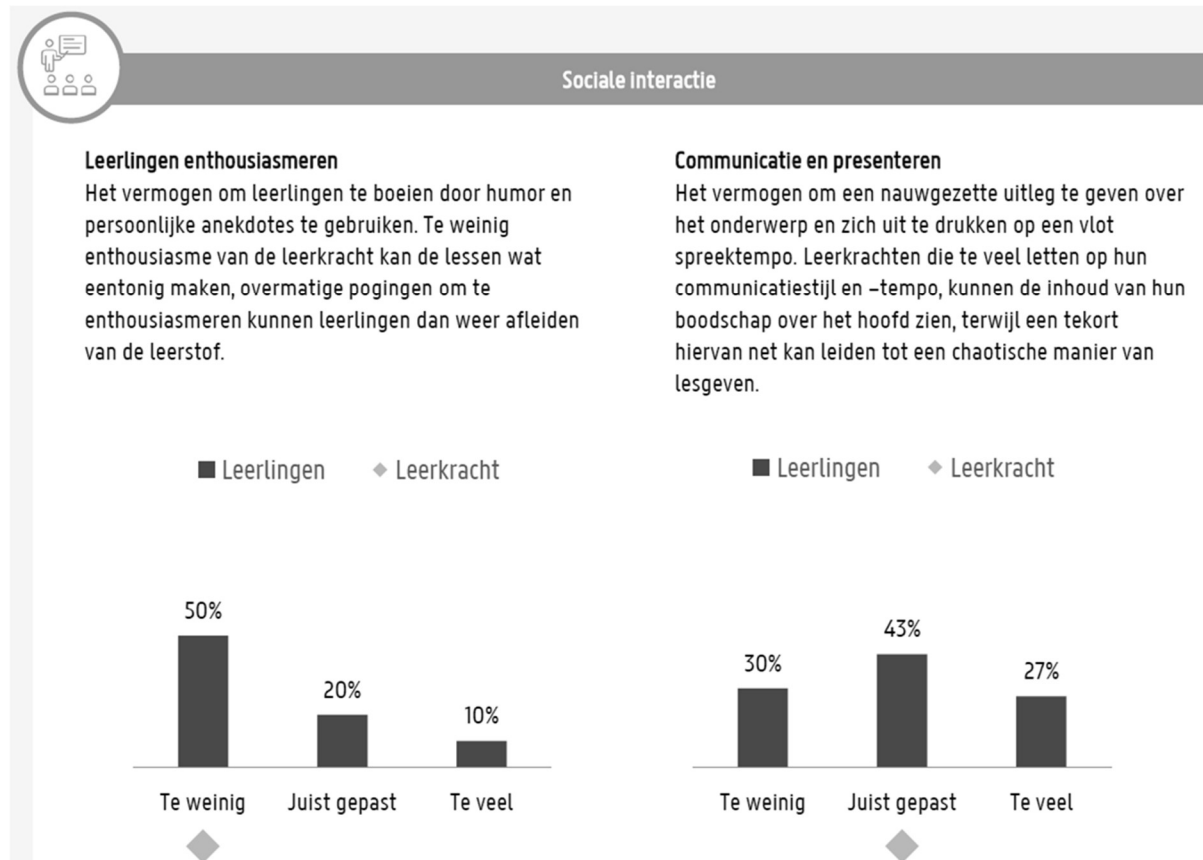
Spider diagram of the current skills (SEMS FFM domains)



Note. SEMS = social and emotional skills; FFM = Five-Factor Model.

Figure A2

Development opportunities for the SEMS domain 'social interaction'



Note. SEMS = social and emotional skills.

Appendix B

Factor Loadings

Table B1 presents the factor loadings of all TLTM items, based on exploratory factor analysis (EFA), conducted separately for teacher and student ratings. Loadings are shown for the full scale, the theoretical domain, and the corresponding facet where applicable. The final subscale structure – comprising seven subscales used in further analyses – is also indicated.

The extraction method was Principal Component Analysis (PCA). Although Varimax rotation was initially planned to help identify uncorrelated components, rotation could not be applied in any of the analyses, as only one component was extracted in each case – whether at the level of the full scale, domains, facets, or subscales.

In factor analysis, the threshold for interpreting item loadings is partly subjective and should be guided by the research context, the desired level of precision, and established statistical conventions. While loadings of .30 are often regarded as the minimum for inclusion, this thesis follows Stevens' (1992) recommendation of using a cut-off of .40 for interpretative purposes, regardless of sample size. Nonetheless, not all loadings in the final subscales reach this threshold. As explained in the Method section, subscale construction was guided by both empirical and theoretical considerations, including internal consistency, explained variance, and the conceptual coherence of item groupings.

Facet-level loadings are only reported when multiple items were available per facet. For single-item facets, no facet loading could be calculated. In the case of two-item facets, identical loadings may appear due to perfect or near-perfect item overlap.

Table B1

TLTM Scale Validation: Factor Loadings for Teachers and Students

TLTM item (with item number)	Teachers				Students			
	Full scale	Dom.	Facet	Final	Full scale	Dom.	Facet	Final
Conscientiousness								
<i>Planning and structuring</i>								
1: Following a lesson plan	-.549	.523	/	.523 ¹	.109	.536	/	.536 ¹
<i>Managing classroom behaviour</i>								
2: Making rules and agreements clear	-.393	.749	.855	.749 ¹	.147	.747	.844	.747 ¹
3: Intervening when rules are broken	-.269	.677	.855	.677 ¹	.020	.732	.844	.732 ¹
<i>Managing attention and participation</i>								
4: Letting the students participate actively	-.228	.548	.835	.548 ¹	.412	.506	.785	.506 ¹
5: Addressing students who lose attention	-.084	.726	.835	.726 ¹	.093	.677	.785	.677 ¹
Agreeableness								
<i>Trusting students</i>								
6: Trusting that students do not cheat	.242	-.070	/	/	.291	.374	/	/
<i>Proactive perspective taking</i>								
7: Paying attention to the needs of students	.069	.564	.825	.688 ²	.638	.697	.841	.754 ²
8: Supporting students who are struggling mentally	.018	.509	.825	.816 ²	.530	.664	.841	.796 ²
<i>Connecting and caring</i>								
9: Building a close bond with students	.375	.592	.867	.867 ³	.524	.581	.804	.804 ³
10: Being involved in students' personal lives	.469	.785	.867	.867 ³	.354	.428	.804	.804 ³
<i>Mindset about students' learning</i>								
11: Believing that every student can get good grades	.111	-.252	.808	.808 ⁴	.372	.554	.837	.837 ⁴
12: Believing that every student can learn	.092	-.377	.808	.808 ⁴	.551	.633	.837	.837 ⁴
Extraversion								
<i>Cheerful engagement</i>								
13: Using personal anecdotes in class	.766	.851	.866	.866 ⁵	.374	.763	.823	.823 ⁵
14: Using humour in class	.544	.770	.866	.866 ⁵	.352	.744	.823	.823 ⁵
<i>Communication and presenting</i>								
15: Paying attention to the pace of speech	-.011	.131	.724	/	.299	.215	.759	/
16: Providing detailed explanation	.452	.659	-.724	/	.508	.508	.759	/
Openness								
<i>Passion for learning and understanding</i>								
17: Questioning things	.367	.638	.823	.650 ⁶	.264	.428	.797	.456 ⁶
18: Exploring a problem	.439	.477	.823	.537 ⁶	.441	.481	.797	.507 ⁶
<i>Creative teaching</i>								
19: Experimenting with new teaching methods	.232	.659	.866	.635 ⁶	.398	.550	.838	.536 ⁶
20: Making teaching materials creative and attractive	.134	.468	.866	.509 ⁶	.510	.620	.838	.619 ⁶
<i>Appreciating diversity</i>								
21: Being open to different opinions	.465	.266	.739	.299 ⁶	.575	.575	.778	.600 ⁶
22: Addressing diversity issues	.232	.534	.739	.494 ⁶	.386	.516	.778	.507 ⁶

Continued on the next page.

Table B1 (Continued)

TLTM item (with item number)	Teachers				Students			
	Full scale	Dom.	Facet	Final	Full scale	Dom.	Facet	Final
Openness (Continued)								
<i>Self-reflection and professional development</i>								
23: Gauging how students experience the lesson	.256	.535	.785	.485 ⁶	.543	.603	.810	.578 ⁶
24: Wanting to improve oneself as a teacher	.256	.498	.785	.532 ⁶	.560	.625	.810	.651 ⁶
<i>Flexibility in teaching</i>								
25: Adapting when the lesson goes differently than expected	.276	.221	.720	.230 ⁶	.454	.491	.769	.494 ⁶
26: Tailoring the teaching method to each student	-.272	.332	.720	.713 ²	.459	.522	.769	.687 ²
Neuroticism								
<i>Stress resilience and recovery</i>								
27: Starting the lesson optimistically	.620	.926	.925	.944 ⁷	.439	.551	.693	.737 ⁷
28: Standing in front of the class relaxed and at ease	.654	.940	.940	.944 ⁷	.456	.700	.798	.745 ⁷
29: Shaking off stressful situations	-.180	.288	.289	/	.288	.578	.578	/
<i>Emotional self-control</i>								
30: Having patience	.156	-.006	.818	/	.463	.653	.800	/
31: Controlling negative emotions	.392	.014	.818	.048 ⁷	.445	.641	.800	.649 ⁷

Note. Domains are displayed in **bold**, facets in *italics*. Each item is a five-point Likert scale ranging from - 2 (too little) to + 2 (too much). After the factor analyses, seven final subscales were chosen, as shown in the last column for both the teachers and the students. The extraction method was PCA. Following Stevens' (1992) recommendation, factor loadings of $\geq .40$ are considered acceptable for interpretative purposes, regardless of sample size. TLTM = Too Little/Too Much; Dom. = Domain; PCA = Principal Component Analysis.

¹ 'Conscientiousness' items: Classroom organisation and management.

² 'Proactive perspective taking' items plus item 26: Student-centred responsiveness.

³ 'Connecting and caring' items: Connecting and caring.

⁴ 'Mindset about students' learning' items: Mindset about students' learning.

⁵ 'Cheerful engagement' items: Cheerful engagement.

⁶ 'Openness' items without item 26: Reflective and innovative teaching.

⁷ 'Neuroticism' items without items 29 and 30: Emotionally grounded presence.

Appendix C

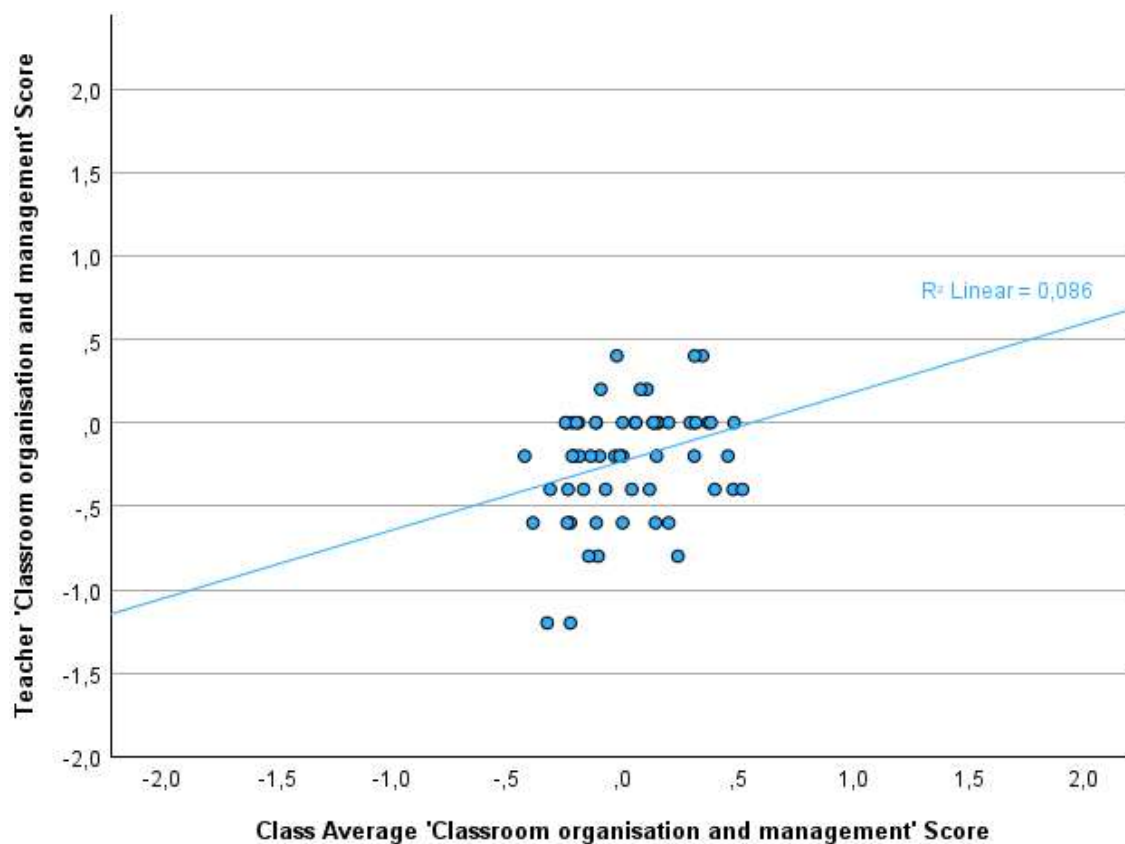
Scatter Plots

In this appendix, scatter plots for all seven subscales are presented, giving a visual representation of the correlation between the teacher score and the class average score for each of the scales. Regression lines are included to facilitate interpretation.

Figure C1

Scatter Plot of Teacher and Class Average 'Classroom organisation and management'

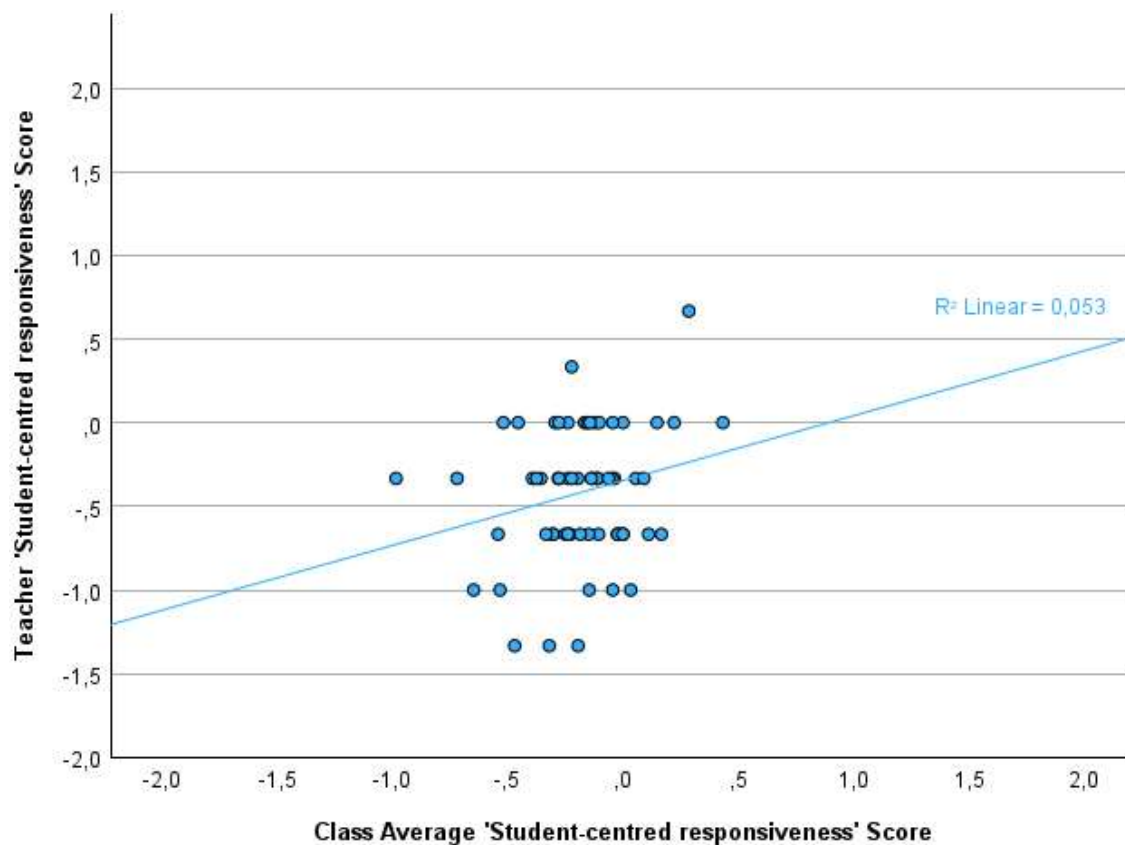
Scores with Regression Line



Note. The x-axis represents the class average of student ratings on the respective TLTM subscale, and the y-axis represents the teacher's self-assessed score on that same subscale. Both axes reflect the full TLTM scale range from -2 (too little) to +2 (too much), with 0 indicating the optimal level of behaviour. The regression line illustrates the strength and direction of the linear association. R^2 = coefficient of determination, with R = Pearson correlation coefficient.

Figure C2

Scatter Plot of Teacher and Class Average 'Student-Centred Responsiveness' Scores with Regression Line

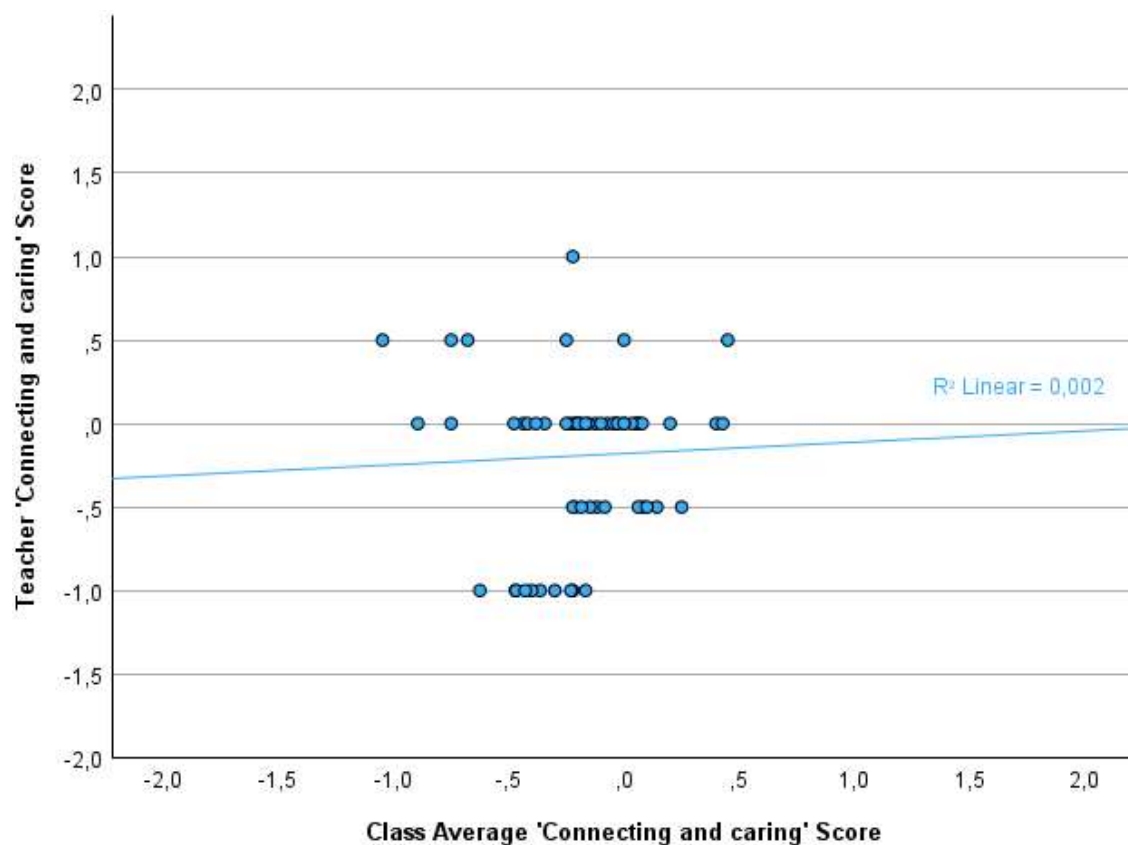


Note. The x-axis represents the class average of student ratings on the respective TLTM subscale, and the y-axis represents the teacher's self-assessed score on that same subscale. Both axes reflect the full TLTM scale range from -2 (too little) to +2 (too much), with 0 indicating the optimal level of behaviour. The regression line illustrates the strength and direction of the linear association. R^2 = coefficient of determination, with R = Pearson correlation coefficient.

Figure C3

Scatter Plot of Teacher and Class Average 'Connecting and Caring' Scores with Regression

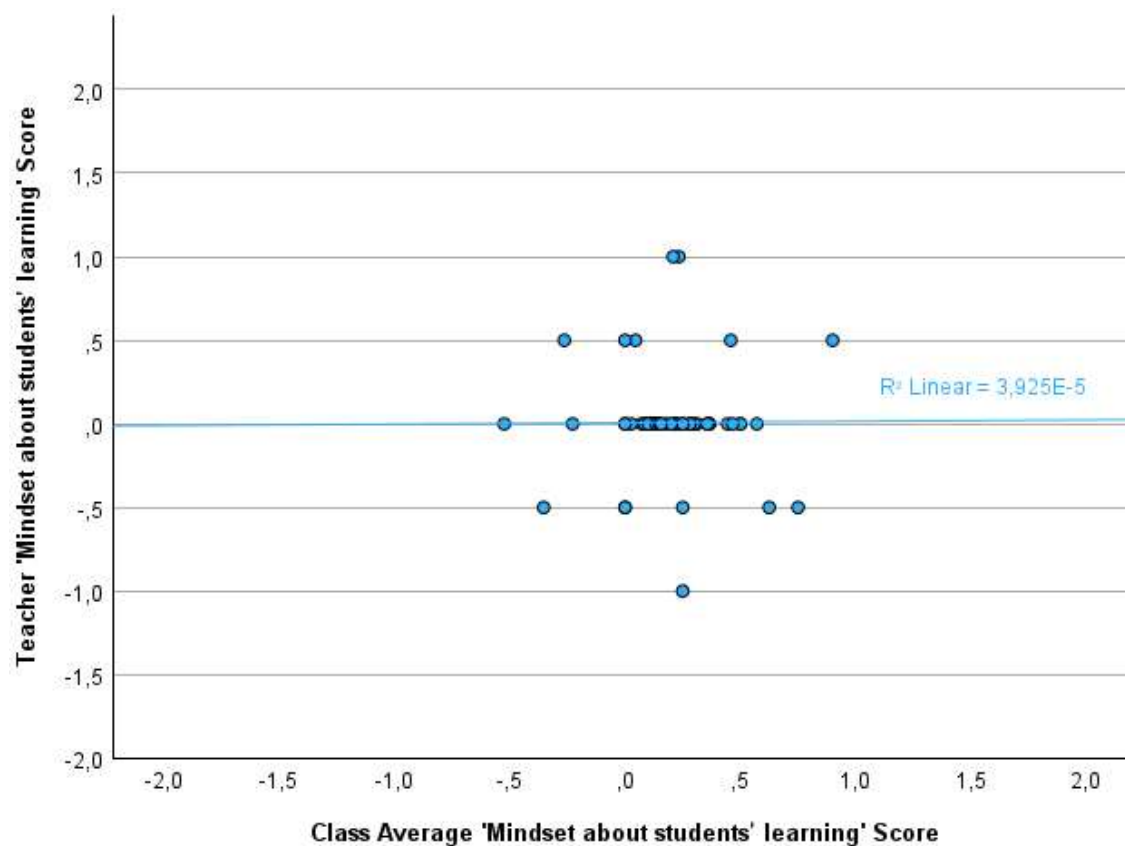
Line



Note. The x-axis represents the class average of student ratings on the respective TLTM subscale, and the y-axis represents the teacher's self-assessed score on that same subscale. Both axes reflect the full TLTM scale range from -2 (too little) to +2 (too much), with 0 indicating the optimal level of behaviour. The regression line illustrates the strength and direction of the linear association. R^2 = coefficient of determination, with R = Pearson correlation coefficient.

Figure C4

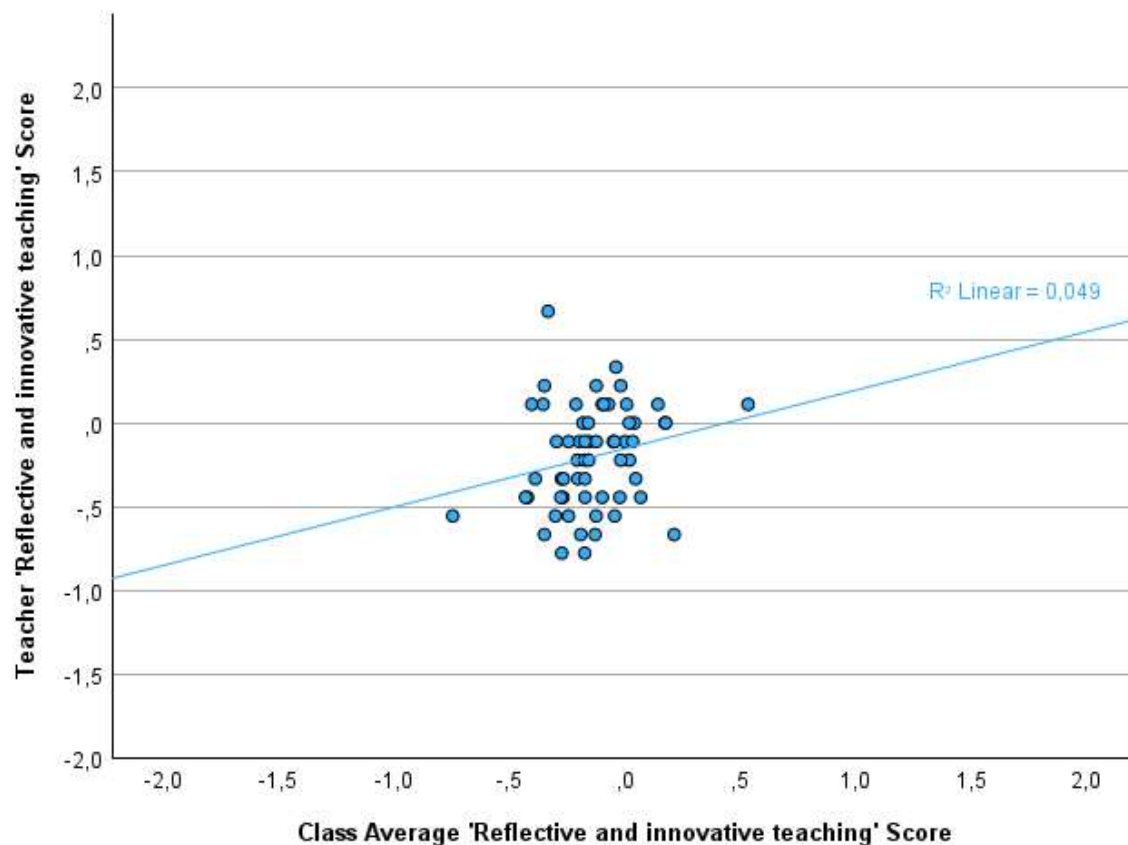
Scatter Plot of Teacher and Class Average 'Mindset About Students' Learning' Scores with Regression Line



Note. The x-axis represents the class average of student ratings on the respective TLTM subscale, and the y-axis represents the teacher's self-assessed score on that same subscale. Both axes reflect the full TLTM scale range from -2 (too little) to +2 (too much), with 0 indicating the optimal level of behaviour. The regression line illustrates the strength and direction of the linear association. R^2 = coefficient of determination, with R = Pearson correlation coefficient.

Figure C6

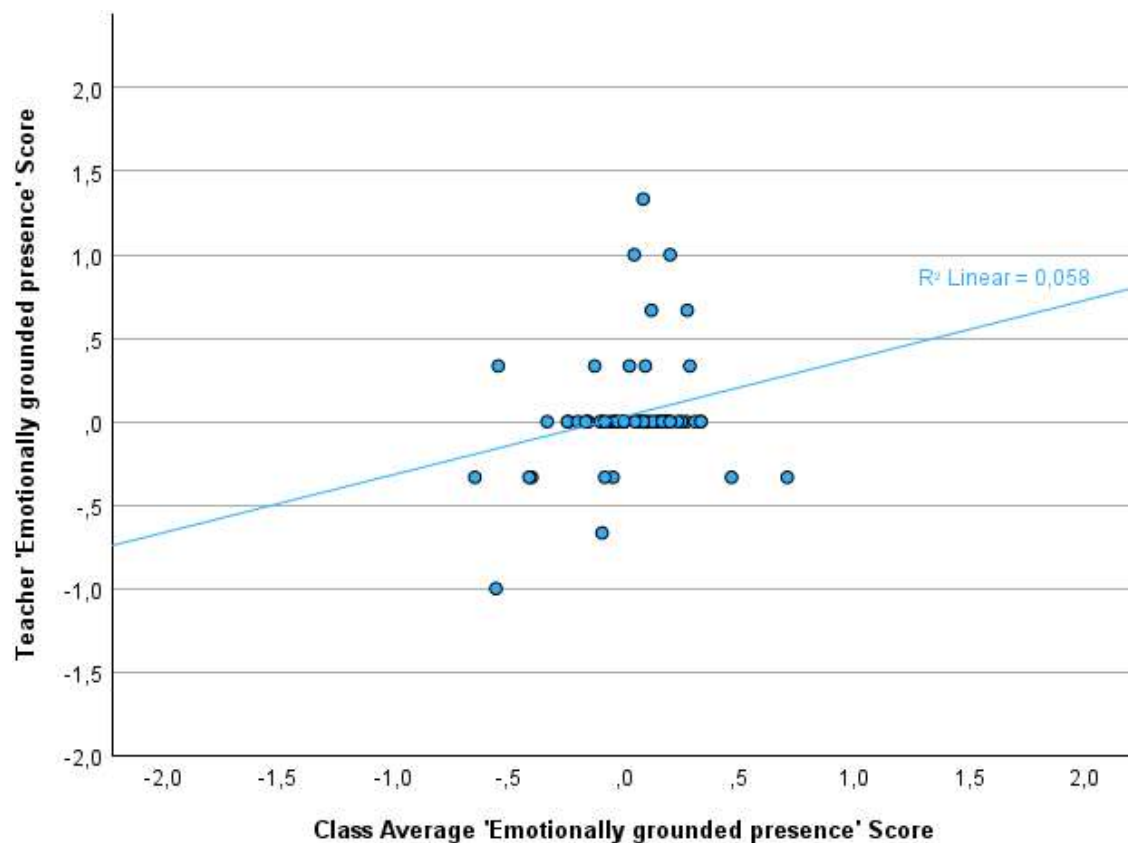
Scatter Plot of Teacher and Class Average 'Reflective and Innovative Teaching' Scores with Regression Line



Note. The x-axis represents the class average of student ratings on the respective TLTM subscale, and the y-axis represents the teacher's self-assessed score on that same subscale. Both axes reflect the full TLTM scale range from -2 (too little) to +2 (too much), with 0 indicating the optimal level of behaviour. The regression line illustrates the strength and direction of the linear association. R^2 = coefficient of determination, with R = Pearson correlation coefficient.

Figure C7

Scatter Plot of Teacher and Class Average 'Emotionally Grounded Presence' Scores with Regression Line



Note. The x-axis represents the class average of student ratings on the respective TLTM subscale, and the y-axis represents the teacher's self-assessed score on that same subscale. Both axes reflect the full TLTM scale range from -2 (too little) to +2 (too much), with 0 indicating the optimal level of behaviour. The regression line illustrates the strength and direction of the linear association. R^2 = coefficient of determination, with R = Pearson correlation coefficient.

Appendix D

Dropout Analysis

Table D1 displays the results of the dropout analysis across the seven Too Little/Too Much (TLTM) subscales. None of the comparisons reached statistical significance after applying the Bonferroni-corrected alpha level of .007, suggesting no substantial differences in T1 scores between teachers who remained in the study and those who dropped out.

Table D1

Mann–Whitney U Test Results for Teacher, Student, and Discrepancy Scores Across the Seven TLTM Subscales

TLTM subscale	Teacher		Students		Discrepancy*	
	z	p	z	p	z	p
Classroom organisation and management	1.519	.129	-0.637	.524	1.807	.071
Student-centred responsiveness	0.852	.394	-1.008	.313	1.412	.158
Connecting and caring	2.054	.040	-1.194	.233	2.138	.033
Mindset about students' learning	-1.220	.222	-0.936	.349	-0.008	.994
Cheerful engagement	0.594	.552	-0.371	.710	1.105	.269
Reflective and innovative teaching	-0.333	.739	-2.000	.045	1.000	.317
Emotionally grounded presence	0.861	.389	-0.896	.371	1.581	.114

Note. As seven comparisons were made per group (Teacher, Student, Discrepancy), a Bonferroni correction was applied; *p*-values should be interpreted against an adjusted alpha level of .007. TLTM = Too Little/Too Much; *z* = standardised test statistic; *p* = *p*-value for statistical significance.

*Teacher self-rating minus class average.